

# Bounding distributional errors via density ratios

LUTZ DÜMBGEN<sup>1</sup>, RICHARD J. SAMWORTH<sup>2</sup>, and JON A. WELLNER<sup>3</sup>

<sup>1</sup>*Department of Mathematics and Statistics, University of Bern, Bern, Switzerland*  
E-mail: [lutz.duembgen@stat.unibe.ch](mailto:lutz.duembgen@stat.unibe.ch)

<sup>2</sup>*Statistical Laboratory, University of Cambridge, Cambridge CB3 0WB, UK*  
E-mail: [r.samworth@statslab.cam.ac.uk](mailto:r.samworth@statslab.cam.ac.uk)

<sup>3</sup>*Statistics, Box 354322, University of Washington, Seattle, WA 98195-4322*  
E-mail: [jaw@stat.washington.edu](mailto:jaw@stat.washington.edu)

We present some new and explicit error bounds for the approximation of distributions. The approximation error is quantified by the maximal density ratio of the distribution  $Q$  to be approximated and its proxy  $P$ . This non-symmetric measure is more informative than and implies bounds for the total variation distance.

Explicit approximation problems include, among others, hypergeometric by binomial distributions, binomial by Poisson distributions, and beta by gamma distributions. In many cases we provide both upper and (matching) lower bounds.

*MSC 2010 subject classifications:* Primary 62E17, Secondary 60E15.

*Keywords:* Binomial distribution, hypergeometric distribution, Poisson approximation, relative errors, total variation distance.

## 1. Introduction

The aim of this work is to provide new inequalities for the approximation of probability distributions. A traditional measure of discrepancy between distributions  $P, Q$  on a space  $(\mathcal{X}, \mathcal{A})$  is their total variation distance

$$d_{\text{TV}}(Q, P) := \sup_{A \in \mathcal{A}} |Q(A) - P(A)|.$$

Alternatively we consider the maximal ratio

$$\rho(Q, P) := \sup_{A \in \mathcal{A}} \frac{Q(A)}{P(A)},$$

with the conventions  $0/0 := 0$  and  $a/0 := \infty$  for  $a > 0$ . Obviously  $\rho(Q, P) \geq 1$  because  $Q(\mathcal{X}) = P(\mathcal{X}) = 1$ . While  $d_{\text{TV}}(\cdot, \cdot)$  is a standard and strong metric on the space of all probability measures on  $(\mathcal{X}, \mathcal{A})$ , the maximal ratio  $\rho(Q, P)$  is particularly important in

situations in which a distribution  $Q$  is approximated by a distribution  $P$ . When  $\rho(Q, P) < \infty$ , we know that

$$Q(A) \leq \rho(Q, P)P(A)$$

for arbitrary events  $A$ , no matter how small  $P(A)$  is, whereas total variation distance gives only the additive bounds  $P(A) \pm d_{\text{TV}}(Q, P)$ .

Explicit values or bounds for  $\rho(Q, P)$  are obtained via density ratios. From now on let  $P$  and  $Q$  have densities  $f$  and  $g$ , respectively, with respect to some measure  $\mu$  on  $(\mathcal{X}, \mathcal{A})$ . Then

$$\rho(Q, P) = \operatorname{ess\,sup}_{x \in \mathcal{X}} \frac{g(x)}{f(x)}. \quad (1)$$

The ratio measure  $\rho(Q, P)$  plays an important role in acceptance-rejection sampling ([von Neumann, 1951](#)): Suppose that  $\rho(Q, P) \leq C < \infty$ . Let  $X_1, X_2, X_3, \dots$  and  $U_1, U_2, U_3, \dots$  be independent random variables where  $X_i \sim P$  and  $U_i \sim \text{Unif}[0, 1]$ . Now let  $\tau_1 < \tau_2 < \tau_3 < \dots$  denote all indices  $i \in \mathbb{N}$  such that  $U_i \leq C^{-1}g(X_i)/f(X_i)$ . Then the random variables  $Y_j := X_{\tau_j}$  and  $W_j := \tau_j - \tau_{j-1}$  ( $j \in \mathbb{N}$ ,  $\tau_0 := 0$ ) are independent with  $Y_j \sim Q$  and  $W_j \sim \text{Geom}(1/C)$ .

As soon as we have a finite bound for  $\rho(Q, P)$ , we can bound total variation distance or other measures of discrepancy. The general result is as follows:

**Proposition 1.** *Suppose that  $g/f \leq \rho$  for some number  $\rho \in [1, \infty)$ .*

*(a) For any non-decreasing function  $\psi : [0, \infty) \rightarrow \mathbb{R}$  with  $\psi(1) = 0$ ,*

$$\int \psi(g/f) dQ \leq Q(\{g > f\})\psi(\rho).$$

*(b) For any convex function  $\psi : [0, \infty) \rightarrow \mathbb{R}$ ,*

$$\int \psi(g/f) dP \leq \psi(0) + \frac{\psi(\rho) - \psi(0)}{\rho}.$$

*Both inequalities are equalities if  $g/f$  takes only values in  $\{0, \rho\}$ .*

Under the assumptions of Proposition 1, the following inequalities hold true, with equality in case of  $g/f \in \{0, \rho\}$ :

**Total variation:** With  $\psi(t) := (1 - t^{-1})_+$ , part (a) leads to

$$d_{\text{TV}}(Q, P) \leq Q(\{g > f\})(1 - \rho^{-1}). \quad (2)$$

**Kullback-Leibler divergence:** With  $\psi(t) := \log t$ , part (a) yields

$$\int \log(g/f) dQ \leq Q(\{g > f\}) \log \rho.$$

**Hellinger distance:** With  $\psi(t) := 2^{-1}(\sqrt{t} - 1)^2$ , part (b) leads to

$$\frac{1}{2} \int (\sqrt{f} - \sqrt{g})^2 d\mu \leq 1 - \rho^{-1/2}.$$

**Pearson  $\chi^2$  divergence:** With  $\psi(t) := (t - 1)^2$ , part (b) yields

$$\int (g/f - 1)^2 dP \leq \rho - 1.$$

Inequality (2) implies that  $d_{\text{TV}}(Q, P) \leq 1 - \rho(Q, P)^{-1}$ , and the latter quantity is easily seen to be the *mixture index of fit* introduced by Rudas et al. (1994),

$$\begin{aligned} \pi^*(P, Q) &:= \min\{\pi \in [0, 1] : P = (1 - \pi)Q + \pi R \text{ for some distribution } R\} \\ &= \min\{\pi \in [0, 1] : P \geq (1 - \pi)Q \text{ on } \mathcal{A}\}. \end{aligned}$$

The remainder of this paper is organized as follows: In Section 2 we present an explicit inequality for  $\rho(Q, P)$  with  $Q$  being a hypergeometric and  $P$  being an approximating binomial distribution. Our result complements results of Diaconis and Freedman (1980), Ehm (1991) and Holmes (2004) for  $d_{\text{TV}}(Q, P)$ .

In Section 3 we first consider the case of  $Q$  being a binomial distribution and  $P$  being the Poisson distribution with the same mean. The corresponding ratio measure  $\rho(Q, P)$  has been analyzed previously by Christensen et al. (1995) and Antonelli and Regoli (2005). Our new explicit bounds bridge the gap between these two works. As a by-product we obtain explicit bounds for  $d_{\text{TV}}(Q, P)$  which are comparable to well-known bounds from the literature. All these bounds carry over to multinomial distributions, to be approximated by a product of Poisson distributions. In particular, we improve and generalize approximation bounds by Diaconis and Freedman (1987). Indeed, at several places we use sufficiency arguments similarly to the latter authors to reduce multivariate approximation problems to univariate ones. Section 4 presents several further examples, most of which are based on approximating beta by gamma distributions.

Most proofs are deferred to Section 5. In particular, we provide a slightly strengthened version of the Stirling–Robbins approximation of factorials (Robbins, 1955) and some properties of the log-gamma function. This part is potentially of independent interest. As notation used throughout, we write  $[a]_0 := 1$  and  $[a]_m := \prod_{i=0}^{m-1} (a - i)$  for real numbers  $a$  and integers  $m \geq 1$ .

## 2. Binomial approximation of hypergeometric distributions

**Sampling from a finite population.** First we revisit a result of Freedman (1977) concerning sampling with and without replacement. For integers  $1 \leq n \leq N$  let  $\mathcal{X} = \{1, \dots, N\}^n$ , the set of all samples of size  $n$  drawn with replacement from  $\{1, \dots, N\}$ . The uniform distribution  $P$  on  $\mathcal{X}$  has weights

$$P(\{x\}) = 1/N^n$$

for  $x = (x_1, \dots, x_n) \in \mathcal{X}$ . When sampling without replacement, we consider the set  $\mathcal{X}_*$  of all  $x$  with all components different, and the distribution  $Q$  with weights

$$Q(\{x\}) = \begin{cases} 1/[N]_n & \text{if } x \in \mathcal{X}_*, \\ 0 & \text{if } x \in \mathcal{X} \setminus \mathcal{X}_*. \end{cases}$$

Consequently,  $dQ/dP = N^n/[N]_n$  on  $\mathcal{X}_*$  and  $dQ/dP = 0$  on  $\mathcal{X} \setminus \mathcal{X}_*$ , so Proposition 1 (a) with  $\psi(t) := (1 - t^{-1})_+$  implies that

$$\rho(Q, P) = N^n/[N]_n \quad \text{and} \quad d_{\text{TV}}(Q, P) = 1 - \rho(Q, P)^{-1} = 1 - [N]_n/N^n. \quad (3)$$

Freedman (1977) showed that

$$1 - \exp\left(-\frac{n(n-1)}{2N}\right) \leq d_{\text{TV}}(Q, P) \leq \frac{n(n-1)}{2N}. \quad (4)$$

Here are two new bounds for  $\rho(Q, P)$  which we will prove in Section 5. The lower bound in the following display follows from Freedman's proof of the lower bound in (4), while the upper bound is new.

$$\frac{n(n-1)}{2N} \leq \log \rho(Q, P) \leq -\frac{n}{2} \log\left(1 - \frac{n-1}{N}\right). \quad (5)$$

From (3) and (4) one would get the upper bound  $-\log(1 - n(n-1)/(2N))$  with the convention that  $\log(t) := -\infty$  for  $t \leq 0$ . For  $n = 2$  this coincides with the upper bound in (5), for  $n \geq 3$  it is strictly larger.

**Hypergeometric and binomial distributions.** Now recall the definition of the hypergeometric distribution: Consider an urn with  $N$  balls,  $L$  of them being black and  $N - L$  being white. Now we draw  $n$  balls at random and define  $X$  to be the number of black balls in this sample. When sampling with replacement,  $X$  has the binomial distribution  $\text{Bin}(n, L/N)$ , and when sampling without replacement ( $n \leq N$ ),  $X$  has the hypergeometric distribution  $\text{Hyp}(N, L, n)$ . Intuitively one would guess that the difference between  $\text{Bin}(n, L/N)$  and  $\text{Hyp}(N, L, n)$  is small when  $n \ll N$ . Note that when Freedman's (1977) result is applied to a particular function, e.g. the number of black balls, the resulting bound is suboptimal because it involves  $n(n-1)/N$  rather than  $n/N$ . Indeed, Diaconis and Freedman (1980) showed that

$$d_{\text{TV}}(\text{Bin}(n, L/N), \text{Hyp}(N, L, n)) \leq 2\frac{n}{N}.$$

Stronger bounds have been obtained by means of the Chen–Stein method. Ehm (1991) showed that with  $p := L/N$ ,

$$\begin{aligned} & d_{\text{TV}}(\text{Hyp}(N, L, n), \text{Bin}(n, L/N)) \\ & \leq \frac{n}{n+1} (1 - p^{n+1} - (1-p)^{n+1}) \frac{n-1}{N-1} \quad \text{for } 1 \leq n \leq \min\{L, N-L\}, \end{aligned} \quad (6)$$

while [Holmes \(2004\)](#) proved that

$$d_{\text{TV}}(\text{Hyp}(N, L, n), \text{Bin}(n, L/N)) \leq \frac{n-1}{N-1}. \quad (7)$$

Our first main result shows that for fixed parameters  $N$  and  $n \leq N/2 + 1$ , the ratio measure  $\rho(\text{Hyp}(N, L, n), \text{Bin}(n, L/N))$  is maximized by  $L = 1$  (and  $L = N - 1$ ):

**Theorem 2.** *For integers  $N, L, n$  with  $1 \leq n \leq N$ ,  $n-1 \leq N/2$  and  $L \in \{0, 1, \dots, N\}$ ,*

$$\begin{aligned} \rho(\text{Hyp}(N, L, n), \text{Bin}(n, L/N)) &\leq \rho(\text{Hyp}(N, 1, n), \text{Bin}(n, 1/N)) \\ &= \left(1 - \frac{1}{N}\right)^{-(n-1)} \\ &\leq \left(1 - \frac{n-1}{N}\right)^{-1}. \end{aligned}$$

Moreover,

$$\begin{aligned} d_{\text{TV}}(\text{Hyp}(N, L, n), \text{Bin}(n, L/N)) &\leq \left(1 - \frac{[L]_n}{[N]_n} - \frac{[N-L]_n}{[N]_n}\right) \left(1 - \left(1 - \frac{1}{N}\right)^{n-1}\right) \\ &\leq \left(1 - \frac{[L]_n}{[N]_n} - \frac{[N-L]_n}{[N]_n}\right) \frac{n-1}{N}. \end{aligned}$$

**Remarks.** Note that our bounds for  $d_{\text{TV}}(\text{Hyp}(N, L, n), \text{Bin}(n, L/N))$  are slightly better than the bound (7) of [Holmes \(2004\)](#). If we fix  $n$  and let  $L, N \rightarrow \infty$  such that  $L/N \rightarrow p \in (0, 1)$ , then our bounds are equal to

$$(1 + o(1))(1 - p^n - q^n) \frac{n-1}{N}$$

and thus similar to the bound (6) of [Ehm \(1991\)](#). If we fix  $L$  and let  $n, N \rightarrow \infty$  such that  $n/N \rightarrow \gamma \in (0, 1)$ , then our two bounds converge to

$$(1 - \gamma^L)(1 - e^{-\gamma}) \leq (1 - \gamma^L)\gamma,$$

whereas the upper bound in (7) tends to  $\gamma$ , and (6) is not applicable.

### 3. Poisson approximations

#### 3.1. Binomial distributions

It is well-known that for  $n \in \mathbb{N}$  and  $p \in [0, 1]$ , the binomial distribution  $\text{Bin}(n, p)$  may be approximated by the Poisson distribution  $\text{Poiss}(np)$  if  $p$  is small. Explicit bounds for the approximation error have been developed in the more general setting of sums of independent but not necessarily identically distributed Bernoulli random variables by

various authors. [Hodges and Le Cam \(1960\)](#) introduced a coupling method which was refined by [Serfling \(1975\)](#) and implies the inequality

$$d_{\text{TV}}(\text{Bin}(n, p), \text{Poiss}(np)) \leq np(1 - e^{-p}) \leq np^2.$$

By direct calculations involving density ratios, [Reiss \(1993\)](#) showed that

$$d_{\text{TV}}(\text{Bin}(n, p), \text{Poiss}(np)) \leq p.$$

Finally, by means of the Chen–Stein method, [Barbour and Hall \(1984\)](#) derived the remarkable bound

$$d_{\text{TV}}(\text{Bin}(n, p), \text{Poiss}(np)) \leq (1 - e^{-np})p. \quad (8)$$

Concerning the ratio measure  $\rho(\text{Bin}(n, p), \text{Poiss}(np))$ , [Christensen et al. \(1995\)](#) showed that

$$\Lambda(p) := \max_{n \geq 1} \log \rho(\text{Bin}(n, p), \text{Poiss}(n, p))$$

is a convex, piecewise linear function of  $p \in [0, 1)$  with  $\lim_{p \rightarrow 1} \Lambda(p) = \infty$  and

$$\Lambda(p) = p \quad \text{for } 0 \leq p \leq \log(2). \quad (9)$$

A close inspection of their proof reveals that  $\Lambda(p)$  is the maximum of the log-ratio measure  $\log \rho(\text{Bin}(n, p), \text{Poiss}(n, p))$  over all integers  $n \leq 1/(1 - p)$ , so the bound  $\Lambda(p)$  is probably rather conservative for large sample sizes  $n$ . Indeed, it follows from the results of [Antonelli and Regoli \(2005\)](#) that for any fixed  $p \in (0, 1)$ ,

$$\lim_{n \rightarrow \infty} \log \rho(\text{Bin}(n, p), \text{Poiss}(np)) = -\log(1 - p)/2 \quad (10)$$

which is substantially smaller than  $\Lambda(p)$ , at least for small values  $p$ . By means of elementary calculations and an appropriate version of Stirling's formula, we shall prove the following bounds:

**Theorem 3.** *For arbitrary  $n \in \mathbb{N}$ ,*

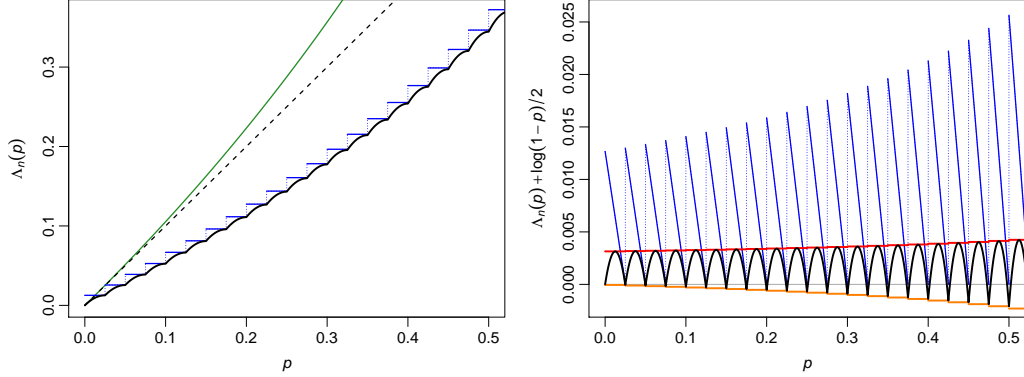
$$\Lambda_n(p) := \log \rho(\text{Bin}(n, p), \text{Poiss}(np))$$

*is a continuous and strictly increasing function of  $p \in [0, 1)$ , satisfying  $\Lambda_n(0) = 0$  and*

$$\Lambda_n(p) < \begin{cases} -\log(1 - p) \\ -\log(1 - \lceil np \rceil/n)/2 \end{cases}$$

*for  $0 < p < 1$ . More precisely, with  $k := \lceil np \rceil$ ,*

$$\Lambda_n(p) + \log(1 - p)/2 \begin{cases} < -\frac{k-1}{12n(n-k+1)} + \frac{1}{8(n-k)+6}, \\ > -\frac{k-1}{12n(n-k+1)} - \frac{1}{12(n-k)(n-k+1)}. \end{cases} \quad (11)$$



**Figure 1.** Comparing  $\text{Bin}(40, p)$  with  $\text{Poiss}(40p)$ .

**Remarks.** Since  $P(\{0\}) = e^{-np} \geq Q(\{0\}) = (1-p)^n$ , the first two upper bounds of Theorem 3 and Proposition 1 (a) lead to the inequalities

$$d_{\text{TV}}(\text{Bin}(n, p), \text{Poiss}(np)) < (1 - (1-p)^n) \cdot \begin{cases} p, \\ 1 - \sqrt{1 - \frac{\lceil np \rceil}{n}} \leq \frac{\lceil np \rceil / n}{2 - \lceil np \rceil / n}; \end{cases}$$

see inequality (20) in Section 5. For fixed  $\lambda > 0$ , the bound in (8) may be rephrased as  $n d_{\text{TV}}(\text{Bin}(n, \lambda/n), \text{Poiss}(\lambda)) \leq (1 - e^{-\lambda})\lambda$ . Our bounds imply that

$$\limsup_{n \rightarrow \infty} d_{\text{TV}}(\text{Bin}(n, \lambda/n), \text{Poiss}(\lambda)) \leq (1 - e^{-\lambda}) \min\{\lambda, \lceil \lambda \rceil / 2\},$$

and  $\lceil \lambda \rceil / 2 < \lambda$  for  $\lambda > 1/2$ . The refined inequalities imply that for any fixed  $p_o \in (0, 1)$ ,

$$\log \rho(\text{Bin}(n, p), \text{Poiss}(np)) = -\log(1-p)/2 + O(n^{-1}) \quad \text{uniformly in } p \leq p_o.$$

The proof of Theorem 3 reveals that  $\Lambda_n(p) = \log \rho(\text{Bin}(n, p), \text{Poiss}(np))$  is concave in  $p \in [(k-1)/n, k/n]$  for each  $k \in \{1, \dots, n\}$ . Figure 1 illustrates this for  $n = 40$ . In the left panel one sees  $\Lambda_n(p)$  (black) together with  $\Lambda(p)$  (black dashed) and the simple upper bounds  $-\log(1-p)$  (green) and  $-\log(1 - \lceil np \rceil / n)/2$  (blue). The right panel shows the quantities  $\Lambda_n(p) + \log(1-p)/2$  (black), i.e. the difference of  $\Lambda_n(p)$  and the asymptotic bound  $-\log(1-p)/2$  of Antonelli and Regoli (2005), together with the upper bound  $-\log(1 - \lceil np \rceil / n)/2 + \log(1-p)/2$  (blue) and the two bounds in (11) (red and orange).

**Poisson binomial distributions.** The distribution  $\text{Bin}(n, p)$  can be replaced with the distribution  $Q$  of  $\sum_{i=1}^n Z_i$  with independent Bernoulli variables  $Z_i$  with arbitrary parameters  $p_i := \mathbb{P}(Z_i = 1) \in (0, 1)$  and  $\lambda := \sum_{i=1}^n p_i$  in place of  $np$ . Dümbgen and Wellner (2020) showed that  $\rho(Q, \text{Poiss}(\lambda)) \leq (1 - p_*)^{-1}$  with  $p_* := \max_{1 \leq i \leq n} p_i$ .

### 3.2. Multinomial distributions and Poissonization

**Multinomial distributions.** The previous bounds for the approximation of binomial by Poisson distributions imply bounds for the approximation of multinomial distributions by products of Poisson distributions. For integers  $n, K \geq 1$  and parameters  $p_1, \dots, p_K > 0$  such that  $p_+ := \sum_{i=1}^K p_i < 1$ , let  $(Y_0, Y_1, \dots, Y_K)$  follow a multinomial distribution

$$\text{Mult}(n; p_0, p_1, \dots, p_K),$$

where  $p_0 := 1 - p_+$ . Further, let  $X_1, \dots, X_K$  be independent Poisson random variables with parameters  $np_1, \dots, np_K$  respectively. Elementary calculations reveal that with  $Y_+ := \sum_{i=1}^K Y_i$  and  $X_+ := \sum_{i=1}^K X_i$ ,

$$\mathcal{L}(Y_1, \dots, Y_K \mid Y_+ = m) = \mathcal{L}(X_1, \dots, X_K \mid X_+ = m) = \text{Mult}\left(m; \frac{p_1}{p_+}, \dots, \frac{p_K}{p_+}\right)$$

for arbitrary integers  $m \geq 0$ . Moreover,

$$Y_+ \sim \text{Bin}(n, p_+) \quad \text{and} \quad X_+ \sim \text{Poiss}(np_+).$$

This implies that for arbitrary integers  $x_1, \dots, x_K \geq 0$  and  $x_+ := \sum_{i=1}^K x_i$ ,

$$\frac{\mathbb{P}(Y_i = x_i \text{ for } 1 \leq i \leq K)}{\mathbb{P}(X_i = x_i \text{ for } 1 \leq i \leq K)} = \frac{\mathbb{P}(Y_+ = x_+)}{\mathbb{P}(X_+ = x_+)}.$$

Consequently, by (1),

$$\rho(\mathcal{L}(X_1, \dots, X_K), \mathcal{L}(Y_1, \dots, Y_K)) = \rho(\text{Bin}(n, p_+), \text{Poiss}(np_+)),$$

and one easily verifies that

$$d_{\text{TV}}(\mathcal{L}(X_1, \dots, X_K), \mathcal{L}(Y_1, \dots, Y_K)) = d_{\text{TV}}(\text{Bin}(n, p_+), \text{Poiss}(np_+)).$$

**Poissonization.** Theorem 3 applies also to Poissonization for empirical processes: Let  $X_1, X_2, X_3, \dots$  be independent random variables with distribution  $P$  on a measurable space  $(\mathcal{X}, \mathcal{A})$ . Let  $M_n$  be the random measure  $\sum_{i=1}^n \delta_{X_i}$ , and let  $\widetilde{M}_n$  be a Poisson process on  $(\mathcal{X}, \mathcal{A})$  with intensity measure  $nP$ . Then  $\widetilde{M}_n$  has the same distribution as  $\sum_{i \leq N_n} \delta_{X_i}$ , where  $N_n \sim \text{Poiss}(n)$  is independent from  $(X_i)_{i \geq 1}$ . For a set  $A_o \in \mathcal{A}$  with  $0 < p_o := P(A_o) < 1$ , the restrictions of the random measures  $M_n$  and  $\widetilde{M}_n$  to  $A_o$  satisfy the equality

$$\rho(\mathcal{L}(M_n|_{A_o}), \mathcal{L}(\widetilde{M}_n|_{A_o})) = \rho(\text{Bin}(n, p_o), \text{Poiss}(np_o)).$$

Here  $M_n|_{A_o}$  and  $\widetilde{M}_n|_{A_o}$  stand for the random measures

$$\{A \in \mathcal{A} : A \subseteq A_o\} \ni A \mapsto M_n(A), \widetilde{M}_n(A)$$



on  $A_o$ . Indeed, for arbitrary integers  $m \geq 0$ ,

$$\mathcal{L}(M_n|_{A_o} \mid M_n(A_o) = m) = \mathcal{L}(\widetilde{M}_n|_{A_o} \mid \widetilde{M}_n(A_o) = m),$$

while

$$M_n(A_o) \sim \text{Bin}(n, p_o) \quad \text{and} \quad \widetilde{M}_n(A_o) \sim \text{Pois}(np_o).$$

Consequently,

$$\rho(\mathcal{L}(M_n|_{A_o}), \mathcal{L}(\widetilde{M}_n|_{A_o})) = \rho(\text{Bin}(n, p_o), \text{Pois}(np_o))$$

and

$$d_{\text{TV}}(\mathcal{L}(M_n|_{A_o}), \mathcal{L}(\widetilde{M}_n|_{A_o})) = d_{\text{TV}}(\text{Bin}(n, p_o), \text{Pois}(np_o)).$$

## 4. Gamma approximations and more

In this section we present further examples of bounds for the ratio measure  $\rho(Q, P)$ . In all but one case, they are related to the approximation of beta by gamma distributions.

### 4.1. Beta distributions

In what follows, let  $\text{Beta}(a, b)$  be the beta distribution with parameters  $a, b > 0$ . The corresponding density is given by

$$\beta_{a,b}(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1} (1-x)_+^{b-1}, \quad x > 0,$$

with the gamma function  $\Gamma(a) := \int_0^\infty x^{a-1} e^{-x} dx$ . Note that we view  $\text{Beta}(a, b)$  as a distribution on the halfline  $(0, \infty)$ , because we want to approximate it by gamma distributions. Specifically, let  $\text{Gamma}(a, c)$  be the gamma distribution with shape parameter  $a > 0$  and rate parameter (i.e. inverse scale parameter)  $c > 0$ . The corresponding density is given by

$$\gamma_{a,c}(x) = \frac{c^a}{\Gamma(a)} x^{a-1} e^{-cx}, \quad x > 0,$$

The next theorem shows that  $\text{Beta}(a, b)$  may be approximated by  $\text{Gamma}(a, c)$  for suitable rate parameters  $c > 0$ , provided that  $b \gg \max(a, 1)$ .

**Theorem 4.** (i) For arbitrary parameters  $a > 0$  and  $b > 1$ ,

$$\begin{aligned} \rho(\text{Beta}(a, b), \text{Gamma}(a, a+b)) &\leq (1-\delta)^{-1/2} \quad \text{and} \\ d_{\text{TV}}(\text{Beta}(a, b), \text{Gamma}(a, a+b)) &\leq 1 - (1-\delta)^{1/2} < \frac{\delta}{2-\delta}, \end{aligned}$$

where

$$\delta := \frac{a+1}{a+b}.$$

(ii) For  $a > 0$ ,  $b > 1$ , and arbitrary  $c > 0$ ,

$$\rho(\text{Beta}(a, b), \text{Gamma}(a, c)) \geq \rho(\text{Beta}(a, b), \text{Gamma}(a, a+b-1)).$$

Moreover, for this opimal rate parameter  $c = a + b - 1$ ,

$$\begin{aligned} \rho(\text{Beta}(a, b), \text{Gamma}(a, a+b-1)) &\leq (1 - \tilde{\delta})^{-1/2} \quad \text{and} \\ d_{\text{TV}}(\text{Beta}(a, b), \text{Gamma}(a, a+b-1)) &\leq 1 - (1 - \tilde{\delta})^{1/2} < \frac{\tilde{\delta}}{2 - \tilde{\delta}}, \end{aligned}$$

where

$$\tilde{\delta} := \frac{a}{a+b-1} < \delta.$$

**Remarks.** The rate parameter  $c = a + b$  is canonical in the sense that the means of  $\text{Beta}(a, b)$  and  $\text{Gamma}(a, a+b)$  are both equal to  $a/(a+b)$ . But note that

$$\frac{\tilde{\delta}}{\delta} = \frac{a}{a+1} \cdot \frac{a+b}{a+b-1} \approx \frac{a}{a+1}$$

if  $b \gg \max\{a, 1\}$ . Hence,  $\text{Gamma}(a, a+b-1)$  yields a remarkably better approximation than  $\text{Gamma}(a, a+b)$ , unless  $a$  is rather large or  $b$  is close to 1.

In the proof of Theorem 4 it is shown that in the special case of  $a = 1$ , one can show the following: For  $b > 1$ ,

$$\log \rho(\text{Beta}(1, b), \text{Gamma}(1, b)) = (b-1) \log(1 - 1/b) + 1,$$

and for  $b \geq 2$ ,

$$\left. \begin{aligned} \log \rho(\text{Beta}(1, b), \text{Gamma}(1, b)) \\ d_{\text{TV}}(\text{Beta}(1, b), \text{Gamma}(1, b)) \end{aligned} \right\} \leq \frac{1}{2b} + \frac{1}{4b^2}.$$

## 4.2. The Lévy–Poincaré projection problem

Let  $\mathbf{U} = (U_1, U_2, \dots, U_n)$  be uniformly distributed on the unit sphere in  $\mathbb{R}^n$ . It is well-known that  $\mathbf{U}$  can be represented as  $\mathbf{Z}/\|\mathbf{Z}\|$  where  $\mathbf{Z} \sim N_n(0, I)$  and  $\|\cdot\|$  denotes standard Euclidean norm. Then the first  $k$  coordinates of  $\mathbf{U}$  satisfy

$$\begin{aligned} \sqrt{n}(U_1, \dots, U_k) &\stackrel{d}{=} (Z_1, \dots, Z_k) / \left( n^{-1} \sum_{j=1}^n Z_j^2 \right)^{1/2} \\ &\rightarrow_d (Z_1, \dots, Z_k) \sim N_k(0, I_k), \end{aligned} \tag{12}$$

since  $n^{-1} \sum_{j=1}^n Z_j^2 \rightarrow_p 1$  by the weak law of large numbers. Indeed, let

$$Q_{n,k} := \mathcal{L}(r_n(U_1, \dots, U_k))$$

with  $r_n > 0$ , and let

$$P_k := \mathcal{L}(Z_1, \dots, Z_k) = N_k(0, I).$$

Diaconis and Freedman (1987) showed that

$$d_{\text{TV}}(Q_{n,k}, P_k) \leq \frac{k+3}{n-k-3} \quad \text{for } 1 \leq k \leq n-4 \text{ and } r_n = \sqrt{n}.$$

By means of Theorem 4, this bound can be improved by a factor larger than 2. The approximation becomes even better if we set  $r_n = \sqrt{n-2}$ . To verify all this, we consider the random variables  $R_k := (\sum_{i=1}^k Z_i^2)$ ,  $R_n := (\sum_{i=1}^n Z_i^2)$  and

$$\mathbf{V} := R_k^{-1}(Z_1, \dots, Z_k).$$

Note that  $\mathbf{V}$  is uniformly distributed on the unit sphere in  $\mathbb{R}^k$  and independent of  $(R_k, R_n)$ . Moreover,

$$(Z_1, \dots, Z_k) = R_k \mathbf{V} \quad \text{and} \quad (U_1, \dots, U_k) = \frac{R_k}{R_n} \mathbf{V}.$$

But  $R_k^2 \sim \text{Gamma}(k/2, 1/2)$  and  $R_k^2/R_n^2 \sim \text{Beta}(k/2, (n-k)/2)$ . Hence,

$$\begin{aligned} \rho(Q_{n,k}, P_k) &= \rho(\mathcal{L}(r_n R_k/R_n), \mathcal{L}(R_k)) \\ &= \rho(\mathcal{L}(R_k^2/R_n^2), \mathcal{L}(r_n^{-2} R_k^2)) \\ &= \rho(\text{Beta}(k/2, (n-k)/2), \text{Gamma}(k/2, r_n^2/2)). \end{aligned}$$

Applying Theorem 4 with  $a := k/2$ ,  $b := (n-k)/2$  and  $c := r_n^2/2$  yields the following bounds:

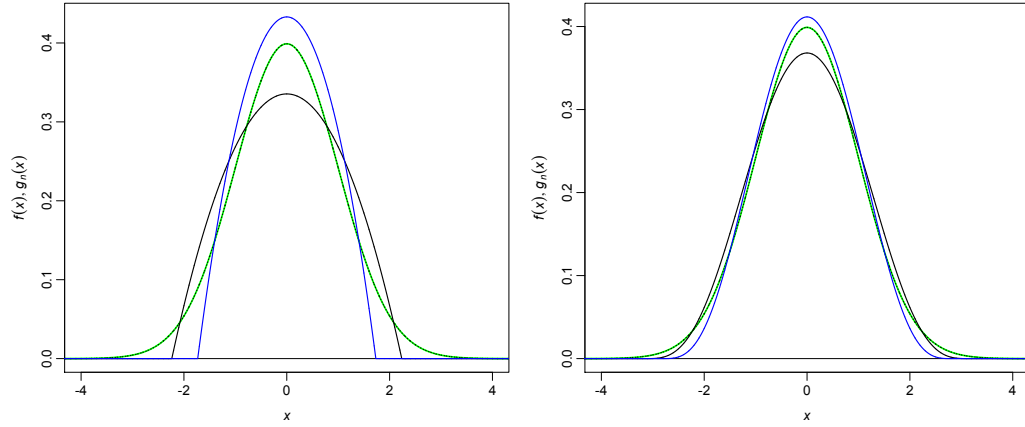
**Corollary 5.** For  $n > k + 2$ ,

$$\begin{aligned} \rho(Q_{n,k}, P_k) &< (1 - \delta)^{-1/2} \quad \text{and} \\ d_{\text{TV}}(Q_{n,k}, P_k) &< 1 - \sqrt{1 - \delta} < \frac{\delta}{2 - \delta}, \end{aligned}$$

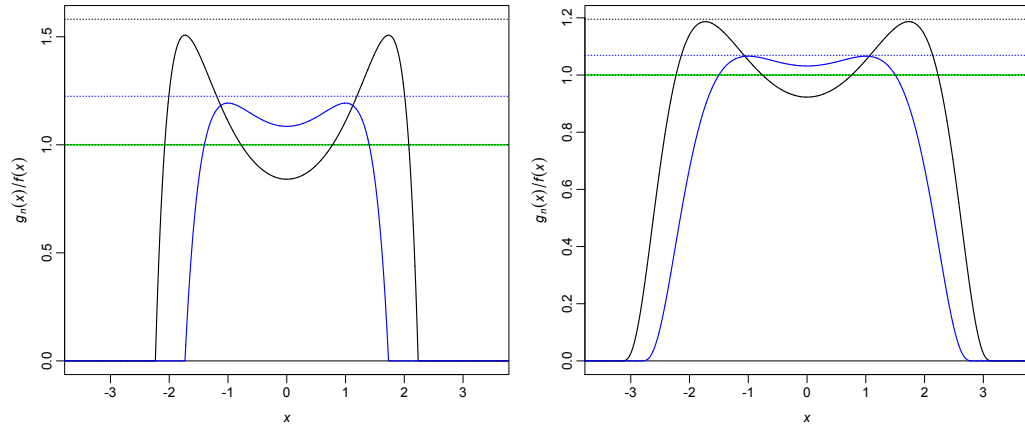
where

$$\delta = \begin{cases} \frac{k+2}{n} & \text{if } r_n = \sqrt{n}, \\ \frac{k}{n-2} & \text{if } r_n = \sqrt{n-2}. \end{cases}$$

Figures 2 and 3 illustrate Corollary 5 in case of  $k = 1$ . For dimensions  $n = 5, 10$ , Figure 2 shows the standard Gaussian density  $f$  (green) and the density  $g_n$  of  $Q_{n,1}$  in case of  $r_n = \sqrt{n}$  (black) and  $r_n = \sqrt{n-2}$  (blue). Figure 3 depicts the corresponding ratios  $g_n/f$ . The dotted black and blue lines are the corresponding upper bounds  $(1 - \delta)^{-1/2}$  from Corollary 5. These pictures show clearly that using  $r_n = \sqrt{n-2}$  instead of  $r_n = \sqrt{n}$  yields a substantial improvement.



**Figure 2.** Densities of  $N(0, 1)$  and  $Q_{n,1}$  for  $n = 5, 10$ .



**Figure 3.** Density ratios for Figure 2.

### 4.3. Dirichlet distributions and uniform spacings

**Dirichlet distributions.** For integers  $1 \leq k \leq N$  and parameters  $a_1, \dots, a_N, c > 0$ , let  $\mathbf{X}$  be a random vector with independent components  $X_i \sim \text{Gamma}(a_i, c)$ . With  $X_+ := \sum_{i=1}^N X_i$ , it is well-known that the random vector

$$\mathbf{Y} = (Y_1, \dots, Y_N) := \left( \frac{X_1}{X_+}, \dots, \frac{X_N}{X_+} \right)$$

and  $X_+$  are independent, where  $X_+ \sim \text{Gamma}(a_+, c)$  with

$$a_+ := \sum_{i=1}^N a_i.$$

The distribution of  $\mathbf{Y}$  is the Dirichlet distribution with parameters  $a_1, \dots, a_N$ , written

$$\mathbf{Y} \sim \text{Dirichlet}(a_1, \dots, a_N).$$

Now let us focus on the first  $k$  components of  $\mathbf{X}$  and  $\mathbf{Y}$ :

$$\begin{aligned} (X_1, \dots, X_k) &= X_+^{(k)}(V_1, \dots, V_k), \\ (Y_1, \dots, Y_k) &= \frac{X_+^{(k)}}{X_+}(V_1, \dots, V_k), \end{aligned}$$

with

$$X_+^{(k)} := \sum_{i=1}^k X_i \quad \text{and} \quad V_i := \frac{X_i}{X_+^{(k)}}.$$

Then  $(V_1, \dots, V_k) \sim \text{Dirichlet}(a_1, \dots, a_k)$  and is independent of  $(X_+^{(k)}, X_+)$ , while

$$\frac{X_+^{(k)}}{X_+} \sim \text{Beta}(a_+^{(k)}, a_+ - a_+^{(k)}) \quad \text{and} \quad X_+^{(k)} \sim \text{Gamma}(a_+^{(k)}, c)$$

with

$$a_+^{(k)} := \sum_{i=1}^k a_i.$$

Hence, the difference between  $\mathcal{L}(Y_1, \dots, Y_k)$  and  $\mathcal{L}(X_1, \dots, X_k)$ , in terms of the ratio measure, is the difference between  $\text{Beta}(a_+^{(k)}, a_+ - a_+^{(k)})$  and  $\text{Gamma}(a_+^{(k)}, c)$ . Thus Theorem 4 yields the following bounds:

**Corollary 6.** *Let  $P_k := \otimes_{i=1}^k \text{Gamma}(a_i, c)$ , and let  $Q_{N,k} := \mathcal{L}(Y_1, \dots, Y_k)$ . Then*

$$\begin{aligned} \rho(Q_{N,k}, P_k) &< (1 - \delta)^{-1/2} \quad \text{and} \\ d_{\text{TV}}(Q_{N,k}, P_k) &< 1 - \sqrt{1 - \delta} < \frac{\delta}{2 - \delta}, \end{aligned}$$

where either

$$c = a_+ \quad \text{and} \quad \delta = \frac{a_+^{(k)} + 1}{a_+},$$

or

$$c = a_+ - 1 \quad \text{and} \quad \delta = \frac{a_+^{(k)}}{a_+ - 1}.$$

**Uniform spacings.** A special case of the previous result are uniform spacings: For an integer  $n \geq 2$ , let  $U_1, \dots, U_n$  be independent random variables with uniform distribution on  $[0, 1]$ . Then we consider the order statistics  $0 < U_{n:1} < U_{n:2} < \dots < U_{n:n} < 1$ . With  $U_{n:0} := 0$  and  $U_{n:n+1} := 1$ , it is well-known that

$$(U_{n:j} - U_{n:j-1})_{j=1}^{n+1} \sim \underbrace{\text{Dirichlet}(1, 1, \dots, 1)}_{n+1 \text{ times}}.$$

That means, the  $n + 1$  spacings have the same distribution as  $(E_j/E_+)_{j=1}^{n+1}$  with independent, standard exponential random variables  $E_1, \dots, E_{n+1}$  and  $E_+ := \sum_{j=1}^{n+1} E_j$ . Consequently, Corollary 6 and the second remark after Theorem 4 yield the following bounds:

**Corollary 7.** For integers  $1 \leq k < n$  let  $Q_{n,k}$  be the distribution of the vector

$$Y_{n,k} := n(U_{n:j} - U_{n:j-1})_{j=1}^k.$$

Further let  $P_k$  be the  $k$ -fold product of the standard exponential distribution. Then

$$\rho(Q_{n,k}, P_k) \leq \begin{cases} \exp\left(\frac{1}{2n} + \frac{1}{4n^2}\right) & \text{if } k = 1, \\ \left(1 - \frac{k}{n}\right)^{-1/2} & \text{in general.} \end{cases}$$

In particular,

$$d_{\text{TV}}(Q_{n,k}, P_k) \leq \begin{cases} \frac{1}{2n} + \frac{1}{4n^2} & \text{if } k = 1, \\ 1 - \sqrt{1 - \frac{k}{n}} < \frac{k}{2n - k} & \text{in general.} \end{cases}$$

**Remarks.** Corollary 7 gives another proof of the results of Runnenburg and Vervaat (1969), who obtained bounds on  $d_{\text{TV}}(Q_{n,k}, P_k)$  by first bounding the Kullback–Leibler divergence; see their Remark 4.1, pages 74–75. It can be shown via the methods of Hall and Wellner (1979) that

$$d_{\text{TV}}(Q_{n,1}, P_1) \leq \frac{2e^{-2}}{n} + \frac{e^{-2}}{n^2},$$

where  $2e^{-2} \approx .2707 < 1/2$ .

#### 4.4. Student distributions

For  $r > 0$  let  $t_r$  denote student's t distribution with  $r$  degrees of freedom, with density

$$f_r(x) = \frac{\Gamma((r+1)/2)}{\Gamma(r/2)\sqrt{r\pi}} \left(1 + \frac{x^2}{r}\right)^{-(r+1)/2}.$$

It is well-known that  $f_r$  converges uniformly to the density  $\phi$  of the standard Gaussian distribution  $N(0, 1)$ , where  $\phi(x) := \exp(-x^2/2)/\sqrt{2\pi}$ . The distribution  $t_r$  has heavier tails than the standard Gaussian distribution and, indeed,

$$\rho(t_r, N(0, 1)) = \infty.$$

However, for the reverse ratio measure we do obtain a reasonable upper bound:

**Lemma 8.** For  $r \geq 2$ ,

$$\frac{1}{2(r+1)} < \log \rho(N(0, 1), t_r) < \frac{1}{2r}.$$

**Remarks.** It follows from Lemma 8 that

$$r \log \rho(N(0, 1), t_r) \rightarrow \frac{1}{2} \quad \text{as } r \rightarrow \infty.$$

By means of Proposition 1 (a) we obtain the inequality  $r d_{\text{TV}}(N(0, 1), t_r) \leq 1/2$  for  $r \geq 2$ . [Pinelis \(2015\)](#) proved that

$$r d_{\text{TV}}(N(0, 1), t_r) < C := \frac{1}{2} \sqrt{\frac{7+5\sqrt{2}}{\pi e^{1+\sqrt{2}}}} \approx 0.3165$$

for  $r \geq 4$ , and that  $r d_{\text{TV}}(N(0, 1), t_r) \rightarrow C$  as  $r \rightarrow \infty$ . So  $C$  is optimal in the bound for  $d_{\text{TV}}$ , whereas  $1/2$  is optimal for  $\rho$ .

Let  $Z$  and  $T_r$  be random variables with distribution  $N(0, 1)$  and  $t_r$ , respectively, where  $r \geq 2$ . Then for any Borel set  $B \subset \mathbb{R}$ ,

$$\mathbb{P}(T_r \in B) \geq e^{-1/(2r)} \mathbb{P}(Z \in B).$$

In particular,

$$\left. \begin{aligned} &\mathbb{P}(\pm T_r < \Phi^{-1}(1 - \alpha)) \\ &\mathbb{P}(|T_r| < \Phi^{-1}(1 - \alpha/2)) \end{aligned} \right\} \geq e^{-1/(2r)}(1 - \alpha).$$

#### 4.5. A counterexample: convergence of normal extremes

In all previous settings, we derived upper bounds for  $\rho(Q, P)$  which implied resonable bounds for  $d_{\text{TV}}(Q, P) = d_{\text{TV}}(P, Q)$ , whereas  $\rho(P, Q) = \infty$  in general. This raises the question whether there are probability densities  $g$  and  $f_n$ ,  $n \geq 1$ , such that  $d_{\text{TV}}(f_n, g) \rightarrow 0$ , but both  $\rho(f_n, g) = \infty$  and  $\rho(g, f_n) = \infty$ ? The answer is “yes” in view of the following example.

**Example 9.** Suppose that  $Z_1, Z_2, Z_3, \dots$  are independent, standard Gaussian random variables. Let  $V_n := \max\{Z_i : 1 \leq i \leq n\}$ . Let  $b_n > 0$  satisfy  $2\pi b_n^2 \exp(b_n^2) = n^2$  and then set  $a_n := 1/b_n$ . Then it is well-known that

$$Y_n := (V_n - b_n)/a_n \rightarrow_d Y_\infty \sim G \quad (13)$$

where  $G$  is the Gumbel distribution function given by  $G(x) = \exp(-\exp(-x))$ . Set  $F_n(x) := P(Y_n \leq x)$  for  $n \geq 1$  and  $x \in \mathbb{R}$ . Hall (1979) shows that for constants  $0 < C_1 < C_2 \leq 3$  and sufficiently large  $n$ ,

$$\frac{C_1}{\log n} < \|F_n - G\|_\infty := \sup_{x \in \mathbb{R}} |F_n(x) - G(x)| < \frac{C_2}{\log n},$$

and  $d_L(F_n, G) = O(1/\log n)$  for the Lévy metric  $d_L$ . It is also known that if  $\tilde{b}_n := (2 \log n)^{1/2} - (1/2)\{\log \log n + \log(4\pi)\}/(2 \log n)^{1/2}$  and  $\tilde{a}_n := 1/\tilde{b}_n$ , then  $\tilde{a}_n/a_n \rightarrow 1$ ,  $(\tilde{b}_n - b_n)/a_n \rightarrow 0$  and (13) continues to hold with  $a_n$  and  $b_n$  replaced by  $\tilde{a}_n$  and  $\tilde{b}_n$ , but the rate of convergence in the last display is not better than  $(\log \log n)^2/\log n$ .

In this example the densities  $f_n$  of  $F_n$  are given by

$$f_n(x) = \Phi(a_n x + b_n)^n \frac{na_n \phi(a_n x + b_n)}{\Phi(a_n x + b_n)} \rightarrow G(x) \cdot e^{-x} = G'(x) =: g(x)$$

for each fixed  $x \in \mathbb{R}$ ; here  $\phi$  is the standard normal density and  $\Phi(z) := \int_{-\infty}^z \phi(y) dy$  is the standard normal distribution function. Thus  $d_{TV}(F_n, G) \rightarrow 0$  by Scheffé's lemma. But in this case it is easily seen that both  $\rho(f_n, g) = \infty$  and  $\rho(g, f_n) = \infty$  where the infinity in the first case occurs in the left tail, and the infinity in the second case occurs in the right tail.

We do not know a rate for the total variation convergence in this example, but it cannot be faster than  $1/\log n$ .

## 5. Proofs and Auxiliary Results

### 5.1. Proofs of the main results

**Proof of (1).** Suppose that  $\mu(\{g/f > r\}) = 0$  for some real number  $r > 0$ . Then  $g \leq rf$ ,  $\mu$ -almost everywhere, so  $Q(A) \leq rP(A)$  for all  $A \in \mathcal{A}$ , and this implies that  $\rho(Q, P) \leq r$ . On the other hand, if  $\mu(\{g/f \geq r\}) > 0$  for some real number  $r > 0$ , then  $A := \{g/f \geq r\} = \{g \geq rf\} \cap \{g > 0\}$  satisfies  $Q(A) > 0$  and  $Q(A) \geq rP(A)$ , whence  $\rho(Q, P) \geq r$ . These considerations show that  $\rho(Q, P)$  equals the  $\mu$ -essential supremum of  $g/f$ .  $\square$

**Proof of Proposition 1.** (a) Under the given hypotheses that  $\psi$  is non-decreasing,  $\psi(1) = 0$  and  $g/f \leq \rho$ , we have

$$\int \psi(g/f) dQ \leq \int_{\{g > f\}} \psi(g/f) dQ \leq Q(\{g > f\})\psi(\rho). \quad (14)$$



Equality holds in the first inequality if and only if  $Q(\{g < f\} \cap \{\psi(g/f) < 0\}) = 0$ , and in the second inequality if and only if  $Q(\{g > f\} \cap \{\psi(g/f) < \psi(\rho)\}) = 0$ . In particular, if  $g/f \in \{0, \rho\}$ , then  $Q(\{g < f\}) = Q(\{g/f = 0\}) = 0$  and  $Q(\{g > f\} \cap \{\psi(g/f) < \psi(\rho)\}) = Q(\emptyset) = 0$ , so we have equality in (14).

(b) For any convex function  $\psi : [0, \infty) \rightarrow \mathbb{R}$  and  $y \in [0, \rho]$ , we have

$$\psi(y) \leq \psi(0) + \frac{y}{\rho} \{\psi(\rho) - \psi(0)\}$$

with equality in case of  $y \in \{0, \rho\}$ . Hence

$$\int \psi(g/f) dP \leq \psi(0) + \frac{\psi(\rho) - \psi(0)}{\rho} \int \frac{g}{f} dP = \psi(0) + \frac{\psi(\rho) - \psi(0)}{\rho}.$$

Equality holds if  $g/f \in \{0, \rho\}$ .  $\square$

**Proof of (5) and comparison with (4).** The asserted bounds are trivial in case of  $n = 1$ , so we assume that  $2 \leq n \leq N$ . Note first that

$$\log \rho(Q, P) = \log(N^n/[N]_n) = \sum_{j=1}^{n-1} H(j)$$

with  $H(x) := -\log(1 - x/N) = \sum_{\ell=1}^{\infty} (x/N)^\ell / \ell$  for  $x \geq 0$ . Since  $H(x) \geq x/N$ ,

$$\log \rho(Q, P) \geq \sum_{j=1}^{n-1} j/N = \frac{n(n-1)}{2N}.$$

This is essentially Freedman's (1977) argument. For the upper bound, it suffices to show that for  $1 \leq n < N$ , the increment

$$\log(N^{n+1}/[N]_{n+1}) - \log(N^n/[N]_n) = H(n) \quad (15)$$

is not larger than the increment

$$-\frac{n+1}{2} \log\left(1 - \frac{n}{N}\right) + \frac{n}{2} \log\left(1 - \frac{n-1}{N}\right) = (n+1)H(n)/2 - nH(n-1)/2. \quad (16)$$

But the difference between (16) and (15) equals

$$(n-1)H(n)/2 - nH(n-1)/2 = n(n-1)(H(n)/n - H(n-1)/(n-1))/2 \geq 0,$$

because  $H(x)/x$  is non-decreasing on  $[0, \infty)$ . Since  $H(tx) > tH(x)$  for  $x \in [0, N)$  and  $t > 1$ , we may also conclude that for  $3 \leq n \leq N$ ,

$$-\log\left(1 - \frac{n(n-1)}{2N}\right) = H(n(n-1)/2) > (n/2)H(n-1) = -\frac{n}{2} \log\left(1 - \frac{n-1}{N}\right).$$

$\square$

**Auxiliary inequalities.** In what follows, we will use repeatedly the following inequalities for logarithms: For real numbers  $x, a > 0$  and  $b > -x$ ,

$$(x+b) \log\left(\frac{x}{x+a}\right) < -a + \frac{a(a-2b)}{2x+a} - \frac{2a^3(x+b)}{3(2x+a)^3} \quad (17)$$

$$< -a + \frac{a(a-2b)}{2x+a} \quad (18)$$

and

$$(x+a/2) \log\left(\frac{x}{x+a}\right) > -a - \frac{a^3}{12x(x+a)}. \quad (19)$$

These inequalities follow essentially from the fact

$$\log\left(\frac{x}{x+a}\right) = \log\left(\frac{2x+a-a}{2x+a+a}\right) = \log\left(\frac{1-y}{1+y}\right) = -2 \sum_{\ell=0}^{\infty} \frac{y^{2\ell+1}}{2\ell+1} < -2y - \frac{2y^3}{3}$$

with  $y := a/(2x+a)$ , where the Taylor series expansion in the second to last step is well-known and follows from the usual expansion  $\log(1 \pm y) = -\sum_{k=1}^{\infty} (\mp y)^k/k$ . Then it follows from  $x+b > 0$  that

$$(x+b) \log\left(\frac{x}{x+a}\right) < -\frac{2a(x+b)}{2x+a} - \frac{2a^3(x+b)}{3(2x+a)^3} = -a + \frac{a(a-2b)}{2x+a} - \frac{2a^3(x+b)}{3(2x+a)^3},$$

whereas

$$\begin{aligned} (x+a/2) \log\left(\frac{x}{x+a}\right) &= \frac{a}{2y} \log\left(\frac{1-y}{1+y}\right) = -a \sum_{\ell=0}^{\infty} \frac{y^{2\ell}}{2\ell+1} \\ &> -a - \frac{ay^2}{3(1-y^2)} = -a - \frac{a^3}{12x(x+a)}. \end{aligned}$$

Here is another expression which will be encountered several times: For  $\delta \in [0, 1]$ ,

$$1 - \sqrt{1-\delta} = \frac{\delta}{1 + \sqrt{1-\delta}} = \frac{\delta}{2 - (1 - \sqrt{1-\delta})} = \dots = \frac{\delta}{2 - \frac{\delta}{2 - \frac{\delta}{2 - \dots}}},$$

and the inequality  $\sqrt{1-\delta} \geq 1 - \delta$  implies that

$$1 - \sqrt{1-\delta} \leq \frac{\delta}{2-\delta} = \frac{\delta}{2} \left(1 - \frac{\delta}{2}\right)^{-1} = \frac{\delta}{2} + \frac{\delta^2}{4-2\delta}. \quad (20)$$

Recall that we write  $[a]_0 := 1$  and  $[a]_m := \prod_{i=0}^{m-1} (a-i)$  for real numbers  $a$  and integers  $m \geq 1$ . In particular,  $\binom{n}{k} = [n]_k/k!$  for integers  $0 \leq k \leq n$ .

**Proof of Theorem 2.** The assertions are trivial in case of  $n = 1$  or  $L \in \{0, N\}$ , because then  $\text{Hyp}(N, L, n) = \text{Bin}(n, L/N)$ . Hence it suffices to consider  $n \geq 2$  and  $1 \leq L \leq N-1$ . For  $k \in \{0, 1, \dots, n\}$  let

$$\begin{aligned} h(k) = h_{N,L,k}(k) &:= \text{Hyp}(N, L, n)(\{k\}) = \binom{L}{k} \binom{N-L}{n-k} / \binom{N}{n} \\ &= \binom{n}{k} \frac{[L]_k [N-L]_{n-k}}{[N]_n}, \\ b(k) = b_{n,L/N}(k) &:= \text{Bin}(n, L/N)(\{k\}) = \binom{n}{k} (L/N)^k (1-L/N)^{n-k} \\ &= \binom{n}{k} \frac{L^k (N-L)^{n-k}}{N^n} \end{aligned}$$

and

$$r(k) = r_{N,L,n}(k) := \frac{h(k)}{b(k)} = \frac{[L]_k [N-L]_{n-k} N^n}{L^k (N-L)^{n-k} [N]_n}.$$

Since

$$r_{N,N-L,n}(n-k) = r_{N,L,n}(k),$$

it even suffices to consider

$$n \geq 2 \quad \text{and} \quad 1 \leq L \leq N/2.$$

In this case,  $r(k) > 0$  for  $1 \leq k \leq \min(n, L)$ , and  $r(k) = 0$  for  $\min(n, L) < k \leq n$ .

In order to maximize the weight ratio  $r$ , note that for any integer  $0 \leq k < \min(L, n)$ ,

$$\frac{r(k+1)}{r(k)} = \frac{(L-k)(N-L)}{L(N-L-n+k+1)} \begin{cases} \leq \\ > \end{cases} 1$$

if and only if

$$k \begin{cases} \geq \\ < \end{cases} \frac{(n-1)L}{N}.$$

Consequently,

$$\begin{aligned} \rho(\text{Hyp}(N, L, n), \text{Bin}(n, L/N)) &= r_{N,L,n}(k) \\ \text{with } k = k_{N,L,n} &:= \left\lceil \frac{(n-1)L}{N} \right\rceil \in \{1, \dots, n-1\}. \end{aligned}$$

The worst-case value  $k_{N,L,n}$  equals 1 if and only if  $L \leq N/(n-1)$ . But

$$\begin{aligned} r_{N,L,n}(1) &= \frac{[N-L]_{n-1} N^n}{(N-L)^{n-1} [N]_n} = \prod_{i=0}^{n-2} \left(1 - \frac{i}{N-L}\right) \frac{N^n}{[N]_n} \\ &\leq \prod_{i=0}^{n-2} \left(1 - \frac{i}{N-1}\right) \frac{N^n}{[N]_n} = (1 - 1/N)^{-(n-1)} = r_{N,1,n}(1). \end{aligned}$$

Consequently, it suffices to consider

$$N/(n-1) < L \leq N/2.$$

Note that these inequalities for  $L$  imply that  $n-1 > 2$ . Hence it remains to prove the assertions when  $n \geq 4$  and  $N/(n-1) < L \leq N/2$ .

The case  $n = 4$  is treated separately: Here it suffices to show that

$$r_{N,L,4}(2) \leq r_{N,1,4}(1) \quad \text{for } N \geq 6 \text{ and } 1 < L \leq N/2.$$

Indeed

$$\begin{aligned} \frac{r_{N,L,4}(2)}{r_{N,1,4}(1)} &= \frac{[L]_2[N-L]_2(N-1)^3}{L^2(N-L)^2[N-1]_3} = \frac{(L-1)(N-L-1)(N-1)^2}{L(N-L)(N-2)(N-3)} \\ &= \frac{(L(N-L) - N + 1)(N-1)^2}{L(N-L)((N-1)^2 - 3N + 5)} = \left(1 - \frac{N-1}{L(N-L)}\right) / \left(1 - \frac{3N-5}{(N-1)^2}\right) \\ &\leq \left(1 - \frac{4(N-1)}{N^2}\right) / \left(1 - \frac{3N-5}{(N-1)^2}\right) \end{aligned}$$

with equality if and only if  $L = N/2$ . The latter expression is less than or equal to 1 if and only if

$$\frac{4(N-1)}{N^2} \geq \frac{3N-5}{(N-1)^2},$$

and elementary manipulations show that this is equivalent to

$$(N - 7/2)^2 + 12 - 49/4 \geq 4/N.$$

But this inequality is satisfied for all  $N \geq 5$ .

Consequently, it suffices to prove our assertion in case of

$$n \geq 5 \quad \text{and} \quad N/(n-1) < L \leq N/2.$$

The maximizer  $k = k_{N,L,n}$  of the density ratio is  $k = \lceil (n-1)L/N \rceil \geq 2$ , and

$$n - k = \lfloor n - (n-1)L/N \rfloor \geq \lfloor n - (n-1)/2 \rfloor = \lfloor (n+1)/2 \rfloor \geq 3.$$

Now our task is to bound

$$\begin{aligned} &\log \rho(\text{Hyp}(N, L, n), \text{Bin}(n, L/N)) \\ &= \log\left(\frac{[L]_k}{L^k}\right) + \log\left(\frac{[N-L]_{n-k}}{(N-L)^{n-k}}\right) - \log\left(\frac{[N]_n}{N^n}\right) \\ &= \log\left(\frac{[L-1]_{k-1}}{L^{k-1}}\right) + \log\left(\frac{[N-L-1]_{n-k-1}}{(N-L)^{n-k-1}}\right) - \log\left(\frac{[N-1]_{n-1}}{N^{n-1}}\right) \end{aligned}$$

from above. Corollary 11 in Section 5.2 implies that for integers  $A \geq m \geq 2$ ,

$$\begin{aligned} \log\left(\frac{[A-1]_{m-1}}{A^{m-1}}\right) &= \log((A-1)!) - \log((A-m)!) - (m-1)\log(A) \\ &= (A-1/2)\log(A) - A - (m-1)\log(A) \\ &\quad - (A-m+1/2)\log(A-m+1) + A-m+1 + s_{m,A} \\ &= (A-m+1/2)\log\left(\frac{A}{A-m+1}\right) + 1-m+s_{m,A}, \end{aligned}$$

where

$$-\frac{m-1}{12A(A-m+1)} < s_{m,A} < 0.$$

Consequently,

$$\begin{aligned} &\log \rho(\text{Hyp}(N, L, n), \text{Bin}(n, L/N)) \\ &< (L-k+1/2)\log\left(\frac{L}{L-k+1}\right) + (N-L-n+k+1/2)\log\left(\frac{N-L}{N-L-n+k+1}\right) \\ &\quad + 1 - (N-n+1/2)\log\left(\frac{N}{N-n+1}\right) + \frac{n-1}{12N(N-n+1)}. \end{aligned}$$

Now we introduce the auxiliary quantities

$$\delta := \frac{n-1}{N}, \quad \Delta := 1 - \delta = \frac{N-n+1}{N}$$

and write

$$k = (n-1)L/N + \gamma = L\delta + \gamma \quad \text{with } 0 \leq \gamma < 1.$$

Then

$$L-k = L\Delta - \gamma, \quad N-L-n+k = (N-L)\Delta + \gamma - 1,$$

whence

$$\begin{aligned} &(L-k+1/2)\log\left(\frac{L}{L-k+1}\right) + (N-L-n+k+1/2)\log\left(\frac{N-L}{N-L-n+k+1}\right) \\ &= (L\Delta + 1/2 - \gamma)\log\left(\frac{L}{L\Delta + 1 - \gamma}\right) + ((N-L)\Delta + \gamma - 1/2)\log\left(\frac{N-L}{(N-L)\Delta + \gamma}\right) \\ &= (L\Delta + 1/2 - \gamma)\log\left(\frac{L\Delta}{L\Delta + 1 - \gamma}\right) + ((N-L)\Delta + \gamma - 1/2)\log\left(\frac{(N-L)\Delta}{(N-L)\Delta + \gamma}\right) \\ &\quad - (N-n+1)\log(\Delta). \end{aligned}$$

It follows from (18) with  $x = L\Delta$ ,  $a = 1 - \gamma$  and  $b = 1/2 - \gamma$  that

$$(L\Delta + 1/2 - \gamma)\log\left(\frac{L\Delta}{L\Delta + 1 - \gamma}\right) < -(1 - \gamma) + \frac{\gamma(1 - \gamma)}{2L\Delta + 1 - \gamma},$$

and with  $x = (N - L)\Delta$ ,  $a = \gamma$  and  $b = \gamma - 1/2$  we may conclude that

$$((N - L)\Delta + \gamma - 1/2) \log\left(\frac{(N - L)\Delta}{(N - L)\Delta + \gamma}\right) < -\gamma + \frac{\gamma(1 - \gamma)}{2(N - L)\Delta + \gamma}.$$

Hence

$$\begin{aligned} & \log \rho(\text{Hyp}(N, L, n), \text{Bin}(n, L/N)) \\ & < - (1 - \gamma) + \frac{\gamma(1 - \gamma)}{2L\Delta + 1 - \gamma} - \gamma + \frac{\gamma(1 - \gamma)}{2(N - L)\Delta + \gamma} - (N - n + 1) \log(\Delta) \\ & \quad + 1 - (N - n + 1/2) \log\left(\frac{N}{N - n + 1}\right) + \frac{n - 1}{12N(N - n + 1)} \\ & = g(L) - \frac{\log(\Delta)}{2} + \frac{\delta}{12N\Delta}, \end{aligned}$$

where

$$\begin{aligned} g(L) &:= \gamma(1 - \gamma) \left( \frac{1}{2L\Delta + 1 - \gamma} + \frac{1}{2(N - L)\Delta + \gamma} \right) \\ &< \frac{1}{8L\Delta} + \frac{1}{8(N - L)\Delta} = \frac{N}{8L(N - L)\Delta}, \end{aligned}$$

because  $\gamma(1 - \gamma) \leq 1/4$ . It will be shown later that

$$g(L) \leq \frac{\delta}{7\Delta}. \quad (21)$$

Consequently,

$$\begin{aligned} \log \rho(\text{Hyp}(N, L, n), \text{Bin}(n, L/N)) &< -\frac{\log(\Delta)}{2} + \frac{\delta}{7\Delta} + \frac{\delta}{12N\Delta} \\ &= -\frac{\log(1 - \delta)}{2} + \frac{\delta}{7(1 - \delta)} + \frac{\delta}{12N(1 - \delta)} \\ &\leq -\frac{\log(1 - \delta)}{2} + \frac{\delta}{7(1 - \delta)} + \frac{\delta}{6N}, \end{aligned}$$

because  $\delta \leq 1/2$ , and we want to show that the right-hand side is not greater than

$$-(n - 1) \log(1 - 1/N) = (n - 1) \sum_{\ell=1}^{\infty} \frac{1}{\ell N^{\ell}} > \delta + \frac{\delta}{2N}.$$

Hence, it suffices to show that

$$-\frac{\log(1 - \delta)}{2} + \frac{\delta}{7(1 - \delta)} - \delta \leq 0.$$

But the left-hand side is a convex function of  $\delta \in [0, 1/2]$  and takes the value 0 for  $\delta = 0$ . Thus it suffices to verify that the latter inequality holds for  $\delta = 1/2$ . Indeed, for  $\delta = 1/2$ , the left-hand side is  $\log(2)/2 + 1/7 - 1/2 = (\log(2) - 5/7)/2 < 0$ .

It remains to verify (21). When  $k = \lceil L\delta \rceil \geq 3$ , this is relatively easy: Here  $2\delta^{-1} < L \leq N/2$ , so

$$L(N - L) > 2\delta^{-1}(N - 2\delta^{-1}) = 2N\delta^{-1}\frac{n-3}{n-1} \geq N\delta^{-1},$$

because  $n \geq 5$ . Hence,

$$g(L) < \frac{N}{8L(N - L)\Delta} < \frac{\delta}{8\Delta}.$$

The case  $k = 2$  is a bit more involved: Since

$$g(L) = \frac{\gamma(1 - \gamma)(2N\Delta + 1)}{(2L\Delta + 1 - \gamma)(2(N - L)\Delta + \gamma)},$$

inequality (21) is equivalent to

$$7\gamma(1 - \gamma)(2N\Delta^2 + \Delta) \leq (2L\Delta + 1 - \gamma)(2(N - L)\Delta + \gamma)\delta. \quad (22)$$

The left-hand side of (22) equals

$$14\gamma(1 - \gamma)N\Delta^2 + 7\gamma(1 - \gamma)\Delta \leq 14\gamma(1 - \gamma)N\Delta^2 + 2\Delta,$$

because  $7\gamma(1 - \gamma) \leq 7/4 < 2$ , while the right-hand of (22) side equals

$$\begin{aligned} 4L(N - L)\Delta^2\delta + 2((1 - \gamma)(N - L) + \gamma L)\Delta\delta + \gamma(1 - \gamma)\delta \\ \geq 4L(N - L)\Delta^2\delta + 2L\delta\Delta > 4L(N - L)\Delta^2\delta + 2\Delta, \end{aligned}$$

because  $N - L \geq L$  and  $L\delta > 1$ . Consequently, it suffices to verify that

$$7\gamma(1 - \gamma)N \leq 2L(N - L)\delta. \quad (23)$$

To this end, note that  $\gamma$  depends on  $L$ , namely,  $\gamma = 2 - L\delta$ , whence  $L = (2 - \gamma)\delta^{-1}$  and

$$2L(N - L)\delta = 2(2 - \gamma)(N - (2 - \gamma)\delta^{-1}) = 2(2 - \gamma)(n - 1 - (2 - \gamma))\delta^{-1},$$

so (23) is equivalent to

$$2(2 - \gamma)(n - 3 + \gamma) - 7\gamma(1 - \gamma)(n - 1) \geq 0. \quad (24)$$

But the left-hand side is

$$\begin{aligned} 4(n - 3) - 2\gamma(4.5n - 8.5) + \gamma^2(7n - 9) \\ \geq 4(n - 3) - \frac{(4.5n - 8.5)^2}{7n - 9} = \frac{4(n - 3)(7n - 9) - (4.5n - 8.5)^2}{7n - 9}. \end{aligned}$$

For  $n \geq 5$ , the denominator is strictly positive, and the derivative of the numerator is  $15.5n - 43.5$ , which is strictly positive, too. Thus it suffices to verify that the numerator is nonnegative for  $n = 5$ . Indeed,  $4(n-3)(7n-9) - (4.5n-8.5)^2 = 12$  for  $n = 5$ .

Finally, it follows from Bernoulli's inequality<sup>1</sup> that  $(1-1/N)^{-(n-1)} \leq (1-(n-1)/N)^{-1}$ . Now the inequalities for the total variation distance are an immediate consequence of Proposition 1 (a) with  $\psi(t) = (1-t^{-1})_+$  and the fact that  $Q(\{0\}) \leq P(\{0\})$  and  $Q(\{n\}) \leq P(\{n\})$ , whence

$$Q(\{g > f\}) \leq 1 - Q(\{0\}) - Q(\{n\}) = 1 - \frac{[N-L]_n}{[N]_n} - \frac{[L]_n}{[N]_n}. \quad \square$$

**Proof of Theorem 3.** Obviously,  $\Lambda_n(0) = 0$ . For  $k \in \mathbb{N}_0$  we introduce the weights  $b(k) = b_{n,p}(k) := \text{Bin}(n, p)(\{k\})$  and  $\pi(k) = \pi_{np}(k) := \text{Poiss}(np)(\{k\}) = e^{-np}(np)^k/k!$ . Obviously,  $b(k) = 0$  for  $k > n$ , while for  $0 \leq k \leq n$  and  $p \in (0, 1)$ ,

$$\lambda_{n,p}(k) := \log \frac{b(k)}{\pi(k)} = \log \left( \frac{[n]_k}{n^k} \right) + np + (n-k) \log(1-p).$$

Note that the right hand side is a continuous function of  $p \in [0, 1)$  with limit  $\lambda_{n,0}(k) := \log([n]_k/n^k) \leq 0$  as  $p \rightarrow 0$ , where  $\lambda_{n,0}(0) = 0$ . Thus we may conclude that

$$\Lambda_n(p) = \max_{k=0,1,\dots,n} \lambda_{n,p}(k)$$

is a continuous function of  $p \in [0, 1)$ .

Next we need to determine the maximizer of  $\lambda_{n,p}(\cdot)$ . For  $k \in \{0, 1, \dots, n-1\}$ ,

$$\lambda_{n,p}(k+1) - \lambda_{n,p}(k) = \log(1-k/n) - \log(1-p) \begin{cases} \geq 0 & \text{if } k \leq np, \\ \leq 0 & \text{if } k \geq np. \end{cases}$$

Consequently,

$$\Lambda_n(p) = \lambda_{n,p}(\lceil np \rceil).$$

From now on we fix an integer  $k \in \{1, \dots, n\}$  and focus on  $p \in [(k-1)/n, k/n]$ , so that  $k = \lceil np \rceil$  if  $p > (k-1)/n$ . Then

$$\Lambda_n(p) = \log \left( \frac{[n]_k}{n^k} \right) + np + (n-k) \log(1-p).$$

This is a concave function of  $p$  with derivative

$$n - \frac{n-k}{1-p} = \frac{k-np}{1-p} \begin{cases} < 1/(1-p) \\ > 0 \end{cases}$$

---

<sup>1</sup> $(1+x)^m \geq 1+mx$  for real numbers  $x > -1$  and  $m \geq 1$



if  $(k-1)/n < p < k/n$ . Since  $1/(1-p)$  is the derivative of  $-\log(1-p)$  with respect to  $p$ , and since  $\Lambda_n(0) = 0 = -\log(1-0)$ , this implies that

$$\Lambda_n(p) < -\log(1-p) \quad \text{for } p \in (0, 1).$$

On the other hand,  $\Lambda_n$  is strictly increasing, whence

$$\Lambda_n(p) \leq \Lambda_n(k/n).$$

But Corollary 11 in Section 5.2 implies that

$$\log\left(\frac{[n]_k}{n^k}\right) = \log\left(\frac{[n-1]_{k-1}}{n^{k-1}}\right) = (n-k+1/2) \log\left(\frac{n}{n-k+1}\right) + 1 - k + s_{k,n}$$

with

$$-\frac{k-1}{12n(n-k+1)} < s_{k,n} < \min\left(0, -\frac{k-1}{12n(n-k+1)} + \frac{1}{12^2(n-k+1)^2}\right).$$

Consequently,

$$\begin{aligned} \Lambda_n(k/n) &= \log\left(\frac{[n]_k}{n^k}\right) + k + (n-k) \log(1-k/n) \\ &\leq (n-k+1/2) \log\left(\frac{n-k}{n-k+1}\right) + 1 - \frac{\log(1-k/n)}{2} \\ &< -\frac{\log(1-k/n)}{2}, \end{aligned}$$

where the last inequality follows from (18) with  $x = n-k$ ,  $a = 1$ , and  $b = 1/2$ .

The refined bounds are for the quantity

$$D_n(p) := \Lambda_n(p) + \log(1-p)/2.$$

For  $p \in [(k-1)/n, k/n]$ ,

$$D_n(p) = \log\left(\frac{[n]_k}{n^k}\right) + np + (n-k+1/2) \log(1-p)$$

and

$$D'_n(p) = n - \frac{n-k+1/2}{1-p} = \frac{k-1/2-np}{1-p} \begin{cases} \geq 0 & \text{if } p \leq (k-1/2)/n, \\ \leq 0 & \text{if } p \geq (k-1/2)/n. \end{cases}$$

Consequently,

$$\begin{aligned} D_n(p) &\leq D_n\left(\frac{k-1/2}{n}\right) \\ &\leq (n-k+1/2) \log\left(\frac{n-k+1/2}{n-k+1}\right) + \frac{1}{2} - \frac{k-1}{12n(n-k+1)} + \frac{1}{12^2(n-k+1)^2}. \end{aligned}$$

It follows from (17) with  $x = n - k + 1/2$ ,  $a = 1/2$  and  $b = 0$  that

$$\begin{aligned} (n - k + 1/2) \log\left(\frac{n - k + 1/2}{n - k + 1}\right) + \frac{1}{2} &= x \log\left(\frac{x}{x + a}\right) + a \\ &< \frac{a^2}{2x + a} - \frac{2a^3x}{3(2x + a)^3} \\ &< \frac{1}{8(n - k) + 6} - \frac{n - k + 1/2}{12 \cdot 8(n - k + 3/4)^3}, \end{aligned}$$

and with  $y := n - k + 3/4 \geq 3/4$ ,

$$\begin{aligned} &\frac{n - k + 1/2}{12 \cdot 8(n - k + 3/4)^3} \Big/ \frac{1}{12^2(n - k + 1)^2} \\ &= \frac{3(y - 1/4)(y + 1/4)^2}{2y^3} > \frac{3(y^2 - 1/16)}{2y^2} \geq \frac{4}{3} \geq 1. \end{aligned}$$

Hence

$$D_n(p) \leq \frac{1}{8(n - k) + 6} - \frac{k - 1}{12n(n - k + 1)}.$$

On the other hand, the lower bound for  $D_n(p)$  in (11) is trivial in case of  $k = n$ , and otherwise

$$\begin{aligned} D_n(p) &\geq \min_{j=k-1, k} D_n(j/n) \\ &= \min_{j=k-1, k} \left( (n - k + 1/2) \log\left(\frac{n - j}{n - k + 1}\right) + 1 - k + j \right) + s_{k,n} \\ &> (n - k + 1/2) \log\left(\frac{n - k}{n - k + 1}\right) + 1 - \frac{k - 1}{12n(n - k + 1)} \\ &> -\frac{1}{12(n - k)(n - k + 1)} - \frac{k - 1}{12n(n - k + 1)} \end{aligned}$$

by (19) with  $x = n - k$  and  $a = 1$ . □

**Proof of Theorem 4.** We start with the first statement of part (ii). Let  $\beta := \beta_{a,b}$  and  $\gamma_c := \gamma_{a,c}$  for  $c > 0$ . Since  $\beta(x) = 0$  for  $x \geq 1$ , it suffices to consider the log-density ratio

$$\lambda_c(x) := \log \frac{\beta}{\gamma_c}(x) = \log \frac{\Gamma(a + b)}{\Gamma(b)} - a \log c + (b - 1) \log(1 - x) + cx$$

for  $0 \leq x < 1$ , noting that the latter expression for  $\lambda_c(x)$  is well-defined for all  $x < 1$ . The derivative of  $\lambda_c$  equals

$$c - \frac{b - 1}{1 - x} = \frac{c}{1 - x} \left( 1 - x - \frac{b - 1}{c} \right) = \frac{c}{1 - x} \left( \frac{c - b + 1}{c} - x \right),$$

and this is smaller or greater than zero if and only if  $x$  is greater or smaller than the ratio  $(c - b + 1)/c$ , respectively. This shows that in case of  $c \leq b - 1$ ,

$$\begin{aligned} \log \rho(\text{Beta}(a, b), \text{Gamma}(a, c)) &= \lambda_c(0) = \log \frac{\Gamma(a+b)}{\Gamma(b)} - a \log c \\ &\geq \log \frac{\Gamma(a+b)}{\Gamma(b)} - a \log(b-1) \\ &= \log \rho(\text{Beta}(a, b), \text{Gamma}(a, b-1)). \end{aligned}$$

For  $c \geq b - 1$ ,

$$\begin{aligned} \log \rho(\text{Beta}(a, b), \text{Gamma}(a, c)) &= \lambda_c\left(\frac{c-b+1}{c}\right) \\ &= \log \frac{\Gamma(a+b)}{\Gamma(b)} - (a+b-1) \log c + (b-1) \log(b-1) + c - b + 1. \end{aligned} \quad (25)$$

But the derivative of the latter expression with respect to  $c \geq b - 1$  equals

$$1 - \frac{a+b-1}{c},$$

so the unique minimizer of  $\log \rho(\text{Beta}(a, b), \text{Gamma}(a, c))$  with respect to  $c > 0$  is  $c = a + b - 1$ .

It remains to verify the inequalities

$$\log \rho(\text{Beta}(a, b), \text{Gamma}(a, a+b)) \leq -\frac{\log(1-\delta)}{2}, \quad (26)$$

$$\log \rho(\text{Beta}(a, b), \text{Gamma}(a, a+b-1)) \leq -\frac{\log(1-\tilde{\delta})}{2}. \quad (27)$$

Then the total variation bounds of Theorem 4 follow from Proposition 1 (a) and the elementary inequality (20). Corollary 11 in Section 5.2 implies that

$$\log \frac{\Gamma(a+b)}{\Gamma(b)} < (a+b-1/2) \log(a+b) - (b-1/2) \log(b) - a. \quad (28)$$

Combining this with (25) yields (26):

$$\begin{aligned} &\log \rho(\text{Beta}(a, b), \text{Gamma}(a, a+b)) \\ &= \log \frac{\Gamma(a+b)}{\Gamma(b)} - (a+b-1) \log(a+b) + (b-1) \log(b-1) + a + 1 \\ &< \frac{\log(a+b)}{2} - \frac{\log(b-1)}{2} + 1 + (b-1/2) \log\left(\frac{b-1}{b}\right) \\ &= -\frac{\log(1-\delta)}{2} + 1 + (b-1/2) \log\left(\frac{b-1}{b}\right) \\ &< -\frac{\log(1-\delta)}{2}, \end{aligned}$$

by (18) with  $(x, a, b) = (b-1, a, 1/2)$ . Concerning (27), it follows from (25) and (28) that

$$\begin{aligned} & \log \rho(\text{Beta}(a, b), \text{Gamma}(a, a+b-1)) \\ &= \log \frac{\Gamma(a+b)}{\Gamma(b)} - (a+b-1) \log(a+b-1) + (b-1) \log(b-1) + a \\ &< \frac{\log(a+b)}{2} - \frac{\log(b-1)}{2} - (a+b-1/2) \log\left(\frac{a+b-1}{a+b}\right) + (b-1/2) \log\left(\frac{b-1}{b}\right) \\ &= -\frac{\log(1-\tilde{\delta})}{2} + \frac{1}{2} \left( A \log\left(\frac{1-1/A}{1+1/A}\right) - B \log\left(\frac{1-1/B}{1+1/B}\right) \right), \end{aligned}$$

where  $A := 2b-1$  and  $B := 2(a+b)-1$ . Now (27) follows from

$$A \log\left(\frac{1-1/A}{1+1/A}\right) - B \log\left(\frac{1-1/B}{1+1/B}\right) = \sum_{\ell=0}^{\infty} \frac{B^{-2\ell} - A^{-2\ell}}{2\ell+1} < 0,$$

because  $A < B$ .

In the special case of  $a = 1$ , we do not need (28) but get via (25) the explicit expression

$$\begin{aligned} \log \rho(\text{Beta}(1, b), \text{Gamma}(1, b)) &= \log \frac{\Gamma(b+1)}{\Gamma(b)} - b \log(b) + (b-1) \log(b-1) + 1 \\ &= (b-1) \log(1-1/b) + 1, \end{aligned}$$

because  $\Gamma(b+1) = b\Gamma(b)$ . Now the standard Taylor series for  $\log(1-x)$  yields that

$$\begin{aligned} & \log \rho(\text{Beta}(1, b), \text{Gamma}(1, b)) \\ &= -(b-1) \sum_{\ell=1}^{\infty} \frac{b^{-\ell}}{\ell} + 1 = \sum_{\ell=1}^{\infty} \left( \frac{b^{-\ell}}{\ell} - \frac{b^{-\ell}}{\ell+1} \right) = \sum_{\ell=1}^{\infty} \frac{b^{-\ell}}{\ell(\ell+1)} \\ &< \frac{1}{2b} + \frac{1}{6b^2} + \frac{1}{12b^3} \sum_{j=0}^{\infty} b^{-j} = \frac{1}{2b} + \frac{1}{6b^2} + \frac{1}{12b^2(b-1)}, \end{aligned}$$

and in case of  $b \geq 2$ , the latter expression is not larger than

$$\frac{1}{2b} + \frac{1}{6b^2} + \frac{1}{12b^2} = \frac{1}{2b} + \frac{1}{4b^2}.$$

□

**Proof of Lemma 8.** By Proposition 1 (a) and the inequality  $1 - \exp(-x) \leq x$  for  $x \geq 0$ , it suffices to verify the claims about  $\log \rho(N(0, 1), t_r)$ . Note first that

$$\log \frac{\phi(x)}{f_r(x)} = \log \frac{\Gamma(r/2) \sqrt{r/2}}{\Gamma((r+1)/2)} + \frac{r+1}{2} \log\left(1 + \frac{x^2}{r}\right) - \frac{x^2}{2}$$

and

$$\frac{\partial}{\partial(x^2)} \log \frac{\phi(x)}{f_r(x)} = \frac{r+1}{2(r+x^2)} - \frac{1}{2} = \frac{1-x^2}{2(r+x^2)},$$

whence

$$\log \rho(N(0, 1), t_r) = \log \frac{\Gamma(r/2)\sqrt{r/2}}{\Gamma((r+1)/2)} - \frac{1}{2} + \frac{r+1}{2} \log\left(1 + \frac{1}{r}\right).$$

On the one hand, the Taylor expansion  $-\log(1-x) = \sum_{k=1}^{\infty} x^k/k$  yields that

$$\begin{aligned} -\frac{1}{2} + \frac{r+1}{2} \log\left(1 + \frac{1}{r}\right) &= -\frac{1}{2} - \frac{r+1}{2} \log\left(\frac{r}{r+1}\right) = -\frac{1}{2} + \frac{r+1}{2} \sum_{k=1}^{\infty} \frac{1}{k(r+1)^k} \\ &= \frac{1}{2} \sum_{k=2}^{\infty} \frac{1}{k(r+1)^{k-1}}, \end{aligned}$$

and the latter series equals

$$\begin{aligned} &\frac{1}{4(r+1)} + \frac{1}{2(r+1)^2} \sum_{\ell=0}^{\infty} \frac{1}{(\ell+3)(r+1)^\ell} \\ &< \frac{1}{4(r+1)} + \frac{1}{6(r+1)^2} \sum_{\ell=0}^{\infty} (r+1)^{-\ell} = \frac{1}{4(r+1)} + \frac{1}{6(r+1)^2(1-(r+1)^{-1})} \\ &= \frac{1}{4(r+1)} + \frac{1}{6(r+1)r} = \frac{1}{4r} - \frac{1}{4r(r+1)} + \frac{1}{6(r+1)r} = \frac{1}{4r} - \frac{1}{12r(r+1)}. \end{aligned}$$

Moreover, it follows from Lemma 12 in Section 5.2 with  $x := r/2$  that

$$\begin{aligned} \log \frac{\Gamma(r/2)\sqrt{r/2}}{\Gamma((r+1)/2)} &< \frac{1}{4r} + \frac{1}{12r(r^2-1)} = \frac{1}{4r} + \frac{1}{12r(r+1)(r-1)} \\ &\leq \frac{1}{4r} + \frac{1}{12r(r+1)}, \end{aligned}$$

because  $r-1 \geq 1$  by assumption. Consequently,

$$\log \rho(N(0, 1), t_r) < \frac{1}{2r}.$$

On the other hand, the previous considerations and Lemma 12 imply that

$$-\frac{1}{2} + \frac{r+1}{2} \log\left(1 + \frac{1}{r}\right) > \frac{1}{4(r+1)}$$

and

$$\log \frac{\Gamma(r/2)\sqrt{r/2}}{\Gamma((r+1)/2)} > \frac{1}{4(r+1)},$$

whence

$$\log \rho(N(0, 1), t_r) > \frac{1}{2(r+1)}.$$

□

## 5.2. Auxiliary Results for the Gamma Function

In what follows, let

$$h(x) := \log \Gamma(x) = \log \int_0^\infty t^{x-1} e^{-t} dt, \quad x > 0.$$

With a random variable  $Y_x \sim \text{Gamma}(x, 1)$  one may write

$$h'(x) = \mathbb{E}(\log Y_x) \quad \text{and} \quad h''(x) = \text{Var}(\log Y_x).$$

The functions  $h'$  and  $h''$  are known as the digamma and trigamma functions; see e.g., Olver et al. (2010), Section 5.15. This shows that  $h(x)$  is strictly convex in  $x > 0$ . Moreover, it follows from concavity of  $\log(\cdot)$  and Jensen's inequality that

$$h'(x) < \log \mathbb{E}(Y_x) = \log x.$$

The well-known identity  $\Gamma(x+1) = x\Gamma(x)$  is equivalent to

$$h(x+1) - h(x) = \log x.$$

**Binet's first formula and Stirling's approximation.** Binet's first integral formula states that

$$h(x) = \log \sqrt{2\pi} + (x - 1/2) \log x - x + R(x), \quad (29)$$

where

$$R(x) := \int_0^\infty e^{-tx} w(t) dt \quad \text{and} \quad w(t) := \frac{1}{t} \left( \frac{1}{2} - \frac{1}{t} + \frac{1}{e^t - 1} \right),$$

see Chapter 12 of [Whittaker and Watson \(1996\)](#). The following lemma provides a lower and upper bound for  $w(t)$ , and these yield rather precise bounds for the remainder  $R(x)$ .

**Lemma 10.** *For arbitrary  $t > 0$ ,*

$$12^{-1} e^{-t/12} < w(t) < 12^{-1}.$$

*In particular, the remainder  $R(x)$  in Binet's formula (29) is strictly decreasing in  $x > 0$  and satisfies*

$$(12x+1)^{-1} < R(x) < (12x)^{-1}.$$

Since  $n! = \Gamma(n+1)$ , Lemma 10 implies a slight improvement of the Stirling approximation by [Robbins \(1955\)](#): For arbitrary integers  $n \geq 0$ ,

$$\log(n!) = \log \sqrt{2\pi} + (n + 1/2) \log(n+1) - n - 1 + s_n \quad (30)$$

with

$$\frac{1}{12(n+1)+1} < s_n < \frac{1}{12(n+1)}.$$

In addition, Binet's formula (29) and Lemma 10 lead to useful inequalities for the increments of  $h(\cdot)$ .

**Corollary 11.** For arbitrary  $0 < a < b$ ,

$$h(b) - h(a) = (b - 1/2) \log(b) - (a - 1/2) \log(a) - (b - a) + s(a, b)$$

where

$$-\frac{b-a}{12ab} < s(a, b) < \min\left(0, -\frac{b-1}{12ab} + \frac{1}{12^2 a^2}\right)$$

**Proof of Lemma 10.** The series expansion of the exponential function and some elementary algebra lead to the representation

$$\begin{aligned} w(t) &= \frac{1}{t} \left( \frac{1}{2} - \frac{1}{t} + \frac{1}{e^t - 1} \right) \\ &= \frac{t(e^t - 1) - 2(e^t - 1 - t)}{2t^3} \bigg/ \frac{e^t - 1}{t} \\ &= \sum_{m=1}^{\infty} \frac{a_m t^{m-1}}{m!} \bigg/ \sum_{m=1}^{\infty} \frac{t^{m-1}}{m!}, \end{aligned}$$

with

$$a_m := \frac{m}{2(m+1)(m+2)}.$$

Note that  $a_1 = 12^{-1}$ , and

$$\frac{a_{m+1}}{a_m} = \frac{(m+1)^2}{(m+1)^2 + m - 1} \begin{cases} = 1 & \text{for } m = 1, \\ < 1 & \text{for } m \geq 2. \end{cases}$$

This shows that  $a_m \leq 12^{-1}$  with strict inequality for  $m \geq 3$ . Consequently,  $w(t) < 12^{-1}$ .

The reverse inequality,  $w(t) > 12^{-1}e^{-t/12}$ , is equivalent to

$$12 \frac{t(e^t - 1) - 2(e^t - 1 - t)}{2t^3} > \frac{e^t - 1}{t} e^{-t/12}.$$

The left hand side equals  $12 \sum_{m=1}^{\infty} a_m t^{m-1}/m!$ , while the right hand side equals

$$\begin{aligned} \frac{e^{(11/12)t} - e^{-t/12}}{t} &= \sum_{m=1}^{\infty} \frac{((11/12)^m - (-1/12)^m) t^{m-1}}{m!} \\ &< \sum_{m=1}^{\infty} \frac{c_m t^{m-1}}{m!} \quad \text{with } c_m := (11/12)^m + (1/12)^m. \end{aligned}$$

Note that  $12a_1 = 1 = c_1$ . Consequently,  $w(t) > 12^{-1}e^{-t/12}$  for all  $t > 0$ , provided that  $12a_m \geq c_m$  for all  $m \geq 2$ . But  $c_2 = 61/72$  and  $c_{m+1}/c_m < 11/12$ , whence  $c_m \leq (61/72)(11/12)^{m-2}$  for  $m \geq 2$ . Consequently, it suffices to show that

$$12(12/11)^{m-2} a_m \geq 61/72 \quad \text{for all } m \geq 2.$$

But

$$\frac{(12/11)^{m+1-2}a_{m+1}}{(12/11)^{m-2}a_m} = \frac{1 + 1/11}{1 + (m-1)/(m+1)^2} < 1$$

if and only if  $m^2 - 9m < -12$ , and for integers  $m \geq 2$  this is equivalent to  $m \leq 7$ . Hence

$$\min_{m \geq 2} 12(12/11)^{m-2}a_m = 12(12/11)^{8-2}a_8 \geq 0.89 > 0.85 \geq 61/72.$$

Since for any fixed  $t > 0$ , the integrand  $e^{-tx}w(t)$  is strictly decreasing in  $x > 0$ , the remainder  $R(x)$  is strictly decreasing in  $x > 0$ . The two bounds for  $w(t)$  imply that  $R(x)$  is larger than  $12^{-1} \int_0^\infty e^{-t(x+1/12)} dt = (12x+1)^{-1}$  and smaller than  $12^{-1} \int_0^\infty e^{-tx} dt = (12x)^{-1}$ .  $\square$

**Proof of Corollary 11.** Writing  $h(x) = \log \sqrt{2\pi} + \tilde{h}(x) + R(x)$  with the auxiliary function  $\tilde{h}(x) := (x - 1/2) \log x - x$ , the remainder term  $s(a, b)$  equals  $R(b) - R(a)$ . But

$$R(a) - R(b) = \int_0^\infty (e^{-ta} - e^{-tb})w(t) dt,$$

and since  $e^{-ta} - e^{-tb} > 0$ , it follows from  $0 < w(t) < 12^{-1}$  that

$$0 < R(a) - R(b) < \frac{1}{12} \int_0^\infty (e^{-ta} - e^{-tb}) dt = \frac{1}{12a} + \frac{1}{12b} = \frac{b-a}{12ab}.$$

Moreover, since  $w(t) > 12^{-1}e^{-t/12}$ ,

$$\begin{aligned} R(a) - R(b) &> \frac{1}{12} \int_0^\infty (e^{-t(a+1/12)} - e^{-t(b+1/12)}) dt = \frac{1}{12a+1} - \frac{1}{12b+1} \\ &= \frac{b-a}{12ab} + \frac{1}{12b(12b+1)} - \frac{1}{12a(12a+1)} > \frac{b-a}{12ab} - \frac{1}{12^2a^2}. \end{aligned} \quad \square$$

**Special increments of  $h$ .** In connection with student distributions, we need lower and upper bounds for the quantities  $h(x+1/2) - h(x) - \log(x)/2$ . With a random variable  $Y_x \sim \text{Gamma}(x, 1)$ , the latter expression equals  $\log \mathbb{E} \sqrt{Y_x/x}$ , so it follows from Jensen's inequality that  $h(x+1/2) - h(x) - \log(x)/2 < \log \sqrt{\mathbb{E}(Y_x/x)} = 0$ . The next lemma shows that  $h(x+1/2) - h(x) - \log(x)/2$  is close to  $-1/(8x)$  for large  $x$ .

**Lemma 12.** For arbitrary  $x > 0$ ,

$$-\frac{1}{8x} - \frac{1}{24x(4x^2-1)_+} < h(x+1/2) - h(x) - \frac{\log x}{2} < -\frac{1}{8(x+1/2)}.$$

**Proof of Lemma 12.** Let us first mention that the second derivative of the log-gamma function  $h$  is given by Gauss' formula

$$h''(x) = \sum_{n=0}^{\infty} \frac{1}{(x+n)^2},$$



see Chapter 12 of [Whittaker and Watson \(1996\)](#). In particular,  $h''$  is strictly convex and decreasing on  $(0, \infty)$  with

$$h''(x) > \int_x^\infty \frac{1}{y^2} dy = \frac{1}{x},$$

because  $(x+n)^{-1} > \int_{x+n}^{x+n+1} y^{-2} dy$ .

Now we start with a general consideration about second order differences of  $h$ : For arbitrary  $0 < a < z$ ,

$$\begin{aligned} h(z+a) + h(z-a) - 2h(z) &= (h(z+a) - h(z)) - (h(z) - h(z-a)) \\ &= \int_0^a (h'(z+u) - h'(z-a+u)) du \\ &= \int_0^a \int_0^a h''(z-a+u+v) dv du \\ &= a^2 \mathbb{E} h''(z-a+a(U+V)), \end{aligned}$$

where  $U$  and  $V$  are independent random variables with uniform distribution on  $[0, 1]$ . Since  $h''$  is convex and  $h''(z) > 1/z$ , it follows from Jensen's inequality that

$$h(z+a) + h(z-a) - 2h(z) \geq a^2 h''(z-a+a\mathbb{E}(U+V)) = a^2 h''(z) > \frac{a^2}{z}.$$

Note also that the distribution of  $W := U+V$  is given by the triangular density  $f(w) := (1-|w-1|)_+$ , so

$$\begin{aligned} h(z+a) + h(z-a) - 2h(z) &= a^2 \int_{\mathbb{R}} (1-|w-1|)_+ h''(z-a+aw) dw \\ &= \int_{\mathbb{R}} (a-|a(w-1)|)_+ h''(z+a(w-1)) a dw \\ &= \int_{\mathbb{R}} (a-|t|)_+ h''(z+t) dt. \end{aligned}$$

We first apply these findings with  $z = x+1/2$  and  $a = 1/2$ : Since  $h(x+1)-h(x) = \log x$ ,

$$\begin{aligned} \frac{\log x}{2} - (h(x+1/2) - h(x)) &= \frac{h(x+1) - h(x)}{2} - h(x+1/2) + h(x) \\ &= \frac{1}{2} (h(x+1) + h(x) - 2h(x+1/2)) \\ &\geq \frac{1}{8(x+1/2)}, \end{aligned}$$

which gives us the upper bound for  $h(x+1/2) - h(x) - \log(x)/2$ . Furthermore,

$$\frac{\log x}{2} - (h(x+1/2) - h(x)) = \frac{1}{2} \int_{\mathbb{R}} (1/2 - |t|)_+ h''(x+1/2+t) dt.$$

On the other hand, if  $x > 1/2$ , then with  $z = x + 1/2$  and  $a = 1$  we obtain

$$\begin{aligned} \log\left(\frac{x+1/2}{x-1/2}\right) &= (h(x+3/2) - h(x+1/2)) - (h(x+1/2) - h(x-1/2)) \\ &= \int_{\mathbb{R}} (1 - |t|)_+ h''(x+1/2+t) dt. \end{aligned}$$

Note that

$$\Delta(t) := \frac{1}{8}(1 - |t|)_+ - \frac{1}{2}(1/2 - |t|)_+$$

has the following properties:

$$\int_{\mathbb{R}} \Delta(t) dt = \int_{\mathbb{R}} \Delta(t)t dt = 0$$

and

$$\Delta(t) \begin{cases} < 0 & \text{if } |t| < 1/3, \\ \geq 0 & \text{if } |t| \geq 1/3. \end{cases}$$

These properties plus the convexity of  $h''$  imply that

$$\int_{\mathbb{R}} \Delta(t) h''(x+1/2+t) dt \geq 0.$$

Indeed, the latter integral doesn't change if we replace  $h''(x+1/2+t)$  with  $g(t) := h''(x+1/2+t) + a + bt$  with constants  $a, b$  such that  $g(\pm 1/3) = 0$ . But then, by convexity of  $g$  and the sign changes of  $\Delta$ , we have that  $g\Delta \geq 0$ . Consequently,

$$\begin{aligned} \frac{\log x}{2} - (h(x+1/2) - h(x)) &= \frac{1}{2} \int_{\mathbb{R}} (1/2 - |t|)_+ h''(x+1/2+t) dt \\ &\leq \frac{1}{8} \int_{\mathbb{R}} (1 - |t|)_+ h''(x+1/2+t) dt \\ &= \frac{1}{8} \log\left(\frac{x+1/2}{x-1/2}\right). \end{aligned}$$

Finally, with  $y := (2x)^{-1} < 1$ , the latter expression equals

$$\begin{aligned} \frac{1}{8} \log\left(\frac{1+y}{1-y}\right) &= \frac{1}{4} \sum_{\ell=0}^{\infty} \frac{y^{2\ell+1}}{2\ell+1} = \frac{y}{4} + \frac{1}{4} \sum_{\ell=1}^{\infty} \frac{y^{2\ell+1}}{2\ell+1} \\ &< \frac{y}{4} + \frac{y^3}{12(1-y^2)} \\ &= \frac{1}{8x} + \frac{1}{24x(4x^2-1)}. \end{aligned} \quad \square$$

**Acknowledgements.** Constructive comments of David Ginsbourger, Dominic Schuhmacher and Kaspar Stucki on an early version of this paper are gratefully acknowledged. We also thank Lutz Mattner for drawing our attention to the technical report [Christensen et al. \(1995\)](#) and for pointing out the connection between the ratio measure and the mixture index of fit. Constructive comments of a referee led to further improvements such as Proposition 1.

This research was supported by Swiss National Science Foundation (L. Dümbgen), an EPSRC fellowship (R. J. Samworth), NSF Grant DMS-1566514 and NI-AID Grant 2R01 AI291968-04 (J. A. Wellner).

## References

- ANTONELLI, S. and REGOLI, G. (2005). On the Poisson-binomial relative error. *Statist. Probab. Lett.* **71** 249–256.
- BARBOUR, A. D. and HALL, P. (1984). On the rate of Poisson convergence. *Math. Proc. Cambridge Philos. Soc.* **95** 473–480.
- CHRISTENSEN, J., FISCHER, P. and KVOLS, K. (1995). On the ratio of binomial and poisson probability distributions. Tech. Rep. 7, Matematisk Institut, Københavns Universitet.
- DIACONIS, P. and FREEDMAN, D. (1980). Finite exchangeable sequences. *Ann. Probab.* **8** 745–764.
- DIACONIS, P. and FREEDMAN, D. (1987). A dozen de Finetti-style results in search of a theory. *Ann. Inst. H. Poincaré Probab. Statist.* **23** 397–423.
- DÜMBGEN, L. and WELLNER, J. A. (2020). The density ratio of Poisson binomial versus Poisson distributions. *Statist. Probab. Lett.* **165** 108862. (arXiv:1910.03444).
- EHM, W. (1991). Binomial approximation to the Poisson binomial distribution. *Statist. Probab. Lett.* **11** 7–16.
- FREEDMAN, D. (1977). A remark on the difference between sampling with and without replacement. *J. Amer. Statist. Assoc.* **72** 681.
- HALL, P. (1979). On the rate of convergence of normal extremes. *J. Appl. Probab.* **16** 433–439.
- HALL, W. J. and WELLNER, J. A. (1979). The rate of convergence in law of the maximum of an exponential sample. *Statist. Neerlandica* **33** 151–154.
- HODGES, J. L., JR. and LE CAM, L. (1960). The Poisson approximation to the Poisson binomial distribution. *Ann. Math. Statist.* **31** 737–740.
- HOLMES, S. (2004). Stein’s method for birth and death chains. In *Stein’s method: expository lectures and applications*, vol. 46 of *IMS Lecture Notes Monogr. Ser.* Inst. Math. Statist., Beachwood, OH, 45–67.
- OLVER, F. W. J., LOZIER, D. W., BOISVERT, R. F. and CLARK, C. W. (eds.) (2010). *NIST handbook of mathematical functions*. U.S. Department of Commerce, National Institute of Standards and Technology, Washington, DC; Cambridge University Press, Cambridge.

- PINELIS, I. (2015). Exact bounds on the closeness between the Student and standard normal distributions. *ESAIM Probab. Stat.* **19** 24–27.
- REISS, R.-D. (1993). *A course on point processes*. Springer Series in Statistics, Springer-Verlag, New York.
- ROBBINS, H. (1955). A remark on Stirling’s formula. *Amer. Math. Monthly* **62** 26–29.
- RUDAS, T., CLOGG, C. C. and LINDSAY, B. G. (1994). A new index of fit based on mixture methods for the analysis of contingency tables. *J. Roy. Statist. Soc. Ser. B* **56** 623–639.
- RUNNENBURG, J. T. and VERVAAT, W. (1969). Asymptotical independence of the lengths of subintervals of a randomly partitioned interval; a sample from S. Ikeda’s work. *Statistica Neerlandica* **23** 67–77.
- SERFLING, R. J. (1975). A general Poisson approximation theorem. *Ann. Probability* **3** 726–731.
- VON NEUMANN, J. (1951). Various techniques used in connection with random digits. monte carlo methods. *Nat. Bureau Standards* **12** 36–38.
- WHITTAKER, E. T. and WATSON, G. N. (1996). *A course of modern analysis*. Cambridge Mathematical Library, Cambridge University Press, Cambridge.